

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:56:02 ; Search time 40.0461 Seconds
(without alignments)
1720.199 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNPVGVGPTSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```

1: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: ./SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	434	23	AAW50080
2	2247	100.0	640	20	AA117090
3	2242	99.8	640	20	AA117091
4	2191	97.5	434	23	AAW50081
5	2183	97.2	639	20	AA117089
6	2155	95.9	639	20	AA117087
7	2143	95.9	640	20	AA117088
8	2143	95.4	434	23	AAW50085
9	2125.5	94.6	433	23	AAW50086
					Bacillus sp KSM-KP
					Bacillus alkaline
					Bacillus alkaline
					Bacillus sp KSM-KP
					Bacillus alkaline
					An alkaline protease
					Bacillus sp alkali
					Bacillus sp alkali

10	2125.5	94.6	641	20	AAW89547
11	2071	92.2	434	23	AAW50090
12	1998.5	88.9	433	23	AAW50084
13	1994.5	88.8	433	23	AAW50082
14	1989.5	88.5	636	20	AAW89548
15	1987.5	88.5	433	23	AAW50083
16	1986.5	88.4	433	13	AAW26274
17	1986.5	88.4	433	19	AAW61495
18	1986.5	88.4	433	20	AAW95698
19	1986.5	88.4	433	21	AAW69207
20	1986.5	88.4	433	21	AAW44619
21	1581.5	70.4	345	19	AAW62230
22	1581.5	70.4	345	20	AAW21654
23	452.5	20.1	659	18	AAW24121
24	452.5	20.1	659	20	AAW94840
25	414	18.4	412	20	AAW94836
26	414	18.4	522	18	AAW24122
27	414	18.4	522	20	AAW94838
28	414	18.4	654	20	AAW24129
29	414	18.4	659	18	AAW94841
30	401	17.8	659	18	AAW24123
31	372	16.6	545	22	ABW09483
32	366	16.3	1079	22	ABW81180
33	366	16.3	1079	24	ABW07391
34	346	15.4	520	18	AAW13666
35	346	15.4	734	18	AAW13667
36	346	15.4	823	18	AAW13668
37	338	15.0	1237	24	ABU11343
38	310.5	13.8	806	13	AAW27481
39	306.5	13.6	519	24	ABP76735
40	306.5	13.6	1938	24	ABP76678
41	304.5	13.6	903	17	AAW87007
42	304.5	13.6	1398	17	AAW87008
43	304.5	13.6	1398	18	AAW24124
44	304.5	13.6	1398	20	AAW94839
45	301.5	13.4	699	20	AAW08471

ALIGNMENTS

RESULT 1

AAW50080

ID AAW50080 standard; protein; 434 AA.

XX AAW50080;

AC AAW50080;

XX AAW50080;

DT 12-AUG-2002 (first entry)

XX Bacillus sp KSM-KP43 alkaline protease protein fragment.

DE Alkaline protease; detergent; laundry; bleaching; dishwasher.

KW Alkaline protease; detergent; laundry; bleaching; dishwasher.

OS Bacillus sp.

XX EPI209233-A2.

XX 29-MAY-2002.

XX 22-NOV-2001; 2001EP-0127851.

XX 22-NOV-2001; 2000JP-0355166.

XX 12-APR-2001; 2001JP-0114048.

XX (KAOS) KAO CORP.

XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;

XX Okuda M, Sacki K;

XX WPI; 2002-437518/47.

XX New modified alkaline proteases useful in detergent compositions -

XX

Bacillus JP170 pro
Bacillus sp KSM-KP
Bacillus sp SD-521
Bacillus sp D6-(FE)
Bacillus sp. alkal
Bacillus sp Y-(FER)
Alkali-protease Ya
Modified Bacillus
Bacillus sp. Lion
Amino acid sequenc
Bacillus Lion Y en
Subtilase JP170 fr
Subtilase JP170.
Thermococcus prote
WO9856926 Seq ID 1
Hyperthermostable
Pyrococcus furiosu
Hyperthermostable
Pyrococcus furiosu
Hyperthermostable
Protease. Synthet
T. yonsei subcili
Transglutaminase r
Foreign protein #4
Fragment of dhpA g
Streptomyces virid
DhpA-mel chimeric
Protein encoded by
RP-III residual pr
Streptomyces virid
Streptomyces virid
Hyperthermostable
Protease. Pyrococ
Pyrococcus furiosu
WO9856926 Seq ID 6
F. balustium CP70

PS Claim 1; Page 10-11; 25pp; English.

XX This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents. The novel proteases have an increased detergency &
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease KP43 from
CC Bacillus sp strain KSM-KP43 which is used to create the modified protease
CC represented in AAM50090.

XX Sequence 434 AA;

Query Match 100.0%; Score 2247; DB 23; Length 434;
Best Local Similarity 100.0%; Pred. No. 3.3e-153;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 60
QY 61 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSDSGGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHNSWGAANGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 121 AGARIHNSWGAANGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPGTFFILSARSSLPDSSF 240
DB 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPGTFFILSARSSLPDSSF 240
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNOGWRVTLDKSLNVAYVNESSLSSTSKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 301 PNGNOGWRVTLDKSLNVAYVNESSLSSTSKATYSFTATAGKPLKISLVWSDAPASTTA 360
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNWDGRNNVNFVINAPOSQGYTIEVOAYN 420
DB 361 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNWDGRNNVNFVINAPOSQGYTIEVOAYN 420
QY 421 VPVGPQTFSLAIVN 434
DB 421 VPVGPQTFSLAIVN 434

RESULT 2

AAV17090
ID AAV17090 standard; Protein; 640 AA.

XX AC AAV17090;
XX 20-MAR-2003 (updated)
DT 21-JUL-1999 (first entry)

XX Bacillus alkaline protease.

XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent.

XX Bacillus sp.

XX WO9918218-A1.

XX 15-APR-1999.

XX 07-OCT-1998; 98WO-JP04528.

XX 07-OCT-1997; 97JP-0274570.

XX

PA (KAOS) KAO CORP.

XX Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
PI Saeiki K, Shikata S, Takaiwa M;
XX WPI; 1999-287736/27.
DR N-PSDB; AAX37278.

PT Alkali protease from Bacillus used in washing powders

XX Disclosure; Page 58-63; 71pp; Japanese.

XX The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease.
CC (Updated on 20-MAR-2003 to correct DR field.)

XX Sequence 640 AA;

Query Match 100.0%; Score 2247; DB 20; Length 640;
Best Local Similarity 100.0%; Pred. No. 5.5e-153;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 60
DB 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITALYALGRTN 266
QY 61 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSDSGGGLGGLPSNLQTLFSQAYS 120
DB 267 NANDTNGHGHVAGSVLGNSTNGKMAPQANLVFQSDSGGGLGGLPSNLQTLFSQAYS 326
QY 121 AGARIHNSWGAANGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
DB 327 AGARIHNSWGAANGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPGTFFILSARSSLPDSSF 240
DB 387 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMPGTFFILSARSSLPDSSF 446
QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB 447 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
QY 301 PNGNOGWRVTLDKSLNVAYVNESSLSSTSKATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 507 PNGNOGWRVTLDKSLNVAYVNESSLSSTSKATYSFTATAGKPLKISLVWSDAPASTTA 566
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNWDGRNNVNFVINAPOSQGYTIEVOAYN 420
DB 567 SVTLVNDLVLITAPNGTQYVGNDFTSFYNDNWDGRNNVNFVINAPOSQGYTIEVOAYN 626
QY 421 VPVGPQTFSLAIVN 434
DB 627 VPVGPQTFSLAIVN 640

RESULT 3

AAV17091
ID AAV17091 standard; Protein; 640 AA.

XX AC AAV17091;

XX 20-MAR-2003 (updated)

DT

Common Applicants

DT XX 21-JUL-1999 (first entry)
DE XX Bacillus alkaline protease.
XX XX
KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
XX washing composition; oxidising agent.
OS XX
XX Bacillus sp.
XX WO9918218-A1.
XX 15-APR-1999.
XX 07-OCT-1998; 98WO-JP04528.
XX 07-OCT-1997; 97JP-0274570.
XX (KAOS) KAO CORP.
XX Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
PI Saeki K, Shikata S, Takaiwa M;
XX WPI; 1999-287736/27.
DR N-PSDB; AAX37279.
XX
XX Alkali protease from Bacillus used in washing powders
XX
XX Disclosure; Page 63-68; 71pp; Japanese.
XX
XX The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease.
CC (Updated on 20-MAR-2003 to correct DR field.)
XX
SQ Sequence 640 AA;

Query Match 99.8%; Score 2242; DB 20; Length 640;
Best Local Similarity 99.8%; Pred. No. 1.3e-152;
Matches 433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60
DB 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 266

QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 267 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 326

QY 121 AGARIHTNSWGAANVCAYTTDSRNVDDYVRKNDMTILFAAGNEGPGGIIISAPGTAKNAI 180
DB 327 AGARIHTNSWGAANVCAYTTDSRNVDDYVRKNDMTILFAAGNEGPGGIIISAPGTAKNAI 386

QY 181 TVGATENLRPSFGSYADNINHVAOFSSRGPTKDGRIKPDVMAPGTILSARSLAPDSSF 240
DB 387 TVGATENLRPSFGSYADNINHVAOFSSRGPTKDGRIKPDVMAPGTILSARSLAPDSSF 446

QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHEVKNRGITPKPSLLKAALIAGAADLGLY 300
DB 447 WANHDSKYAYMGGTSMATPIVAGNVAQLREHEVKNRGITPKPSLLKAALIAGAADLGLY 506

QY 301 PNGOGWGRVTLDKSLNVAAYNESSLSSTSQATYSFTATAGKPLKISLVWSDAPASTTA 360
DB 507 PNGOGWGRVTLDKSLNVAAYNESSLSSTSQATYSFTATAGKPLKISLVWSDAPASTTA 566

QY 361 SVTLVNDLDELVTAPNGTQYVGNDFTSYNDNMWDGRNNVNFVINAQSGTYYIEVOAYN 420
DB 567 SVTLVNDLDELVTAPNGTQYVGNDFTSYNDNMWDGRNNVNFVINAQSGTYYIEVOAYN 626

QY 421 VPVGPFQTFSLAIVN 434
DB 627 VPVGPFQTFSLAIVN 640

RESULT 4

AAM50081
ID AAM50081 standard; protein; 434 AA.
XX

AC AAM50081;
XX

DT 12-AUG-2002 (first entry)
XX

DE Bacillus sp KSM-KP9860 alkaline protease protein fragment.
XX

KW Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX

OS Bacillus sp.
XX

PN EP1209233-A2.
XX

PD 29-MAY-2002.
XX

PF 22-NOV-2001; 2001EP-0127851.
XX

PR 22-NOV-2000; 2000JP-0355166.
XX

PR 12-APR-2001; 2001JP-0114048.
XX

PA (KAOS) KAO CORP.
XX

PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
PI Okuda M, Saeki K;
XX

DR WPI; 2002-437518/47.
XX

PT New modified alkaline proteases useful in detergent compositions -
XX

PS Claim 5; Page 12-13; 25pp; English.
XX

CC This invention describes novel Bacillus sp. alkaline proteases useful in
CC detergent compositions, especially in laundry, bleaching or automatic
CC dishwasher detergents. The novel proteases have an increased detergency &
CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC sequence represents a fragment of the alkaline protease KP9860 from
CC Bacillus sp strain KSM-KP9860 described in the method of the invention.
XX

SQ Sequence 434 AA;

Query Match 97.5%; Score 2191; DB 23; Length 434;
Best Local Similarity 96.5%; Pred. No. 3.5e-149;
Matches 419; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60
DB 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRN 60

QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120

QY 121 AGARIHTNSWGAANVCAYTTDSRNVDDYVRKNDMTILFAAGNEGPGGIIISAPGTAKNAI 180
DB 121 AGARIHTNSWGAANVCAYTTDSRNVDDYVRKNDMTILFAAGNEGPGGIIISAPGTAKNAI 180

QY 181 TVGATENLRPSFGSYADNINHVAOFSSRGPTKDGRIKPDVMAPGTILSARSLAPDSSF 240
DB 181 TVGATENLRPSFGSYADNINHVAOFSSRGPTKDGRIKPDVMAPGTILSARSLAPDSSF 240

241 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLY 300
|||||
241 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADVGUGY 300
|||||
301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYFTATAGKPLKISLVNSDAPASTTA 360
|||||
301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYFTATAGKPLKISLVNSDAPASTTA 360
|||||
361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNWDGRNNVNFVINAPOSQGTYYTIEVOAYN 420
|||||
361 SVTLVNDLVLITAPNGTRYVGNDFTSAPFNNWDGRNNVNFVINSPOSQGTYYTIEVOAYN 420
|||||
421 VPVGPQTFSLAIVN 434
|||||
421 VPVGPQNFSLAIVN 434
|||||

RESULT 5
AAY17089
ID AAY17089 standard; Protein: 639 AA.
XX AAY17089;
XX
XX
XX
DT 20-MAR-2003 (updated)
DT 21-JUL-1999 (first entry)
XX
XX
DE Bacillus alkaline protease.
XX
XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent.
XX
XX Bacillus sp.
XX
XX
XX WO9918218-A1.
XX
XX 15-APR-1999.
XX
XX 07-OCT-1998; 98WO-JP04528.
XX
XX 07-OCT-1997; 97JP-0274570.
XX
XX (KAOS) KAO CORP.
XX
XX Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
PI Saeki K, Shikata S, Takaiwa M;
XX
XX WPI; 1999-287736/27.
XX
XX N-PSDB; AAX37277.
XX
XX Alkali protease from Bacillus used in washing powders
XX
XX Disclosure; Page 53-58; 71pp; Japanese.
XX
XX The invention relates to alkaline proteases produced by strains of
CC Bacillus. The proteases ability to digest casein is not inhibited by
CC oleic acid and they have a high stability to oxidising agents. The
CC alkaline protease of the invention has the following properties: (a) it
CC is active over the pH range 4-13 and has at least 80% of its optimum
CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
CC its ability to digest casein is not inhibited by oleic acid; (e) it has
CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
CC used as enzymes in washing compositions for use in automatic dishwashers
CC and for washing clothes. The stability to oxidising agents allows the
CC enzyme to be an effective component of washing compositions including
CC bleaches. The present sequence represents an alkaline protease.
XX (Updated on 20-MAR-2003 to correct DR field.)
XX
XX Sequence 639 AA;
SQ

Query Match 97.2%; Score 2183; DB 20; Length 639;
Best Local Similarity 96.3%; Pred. No. 2.2e-148;
Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAOSSYGLYGOGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
|||||
Db 206 NDVARGIVKADVAOSSYGLYGOGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 265
|||||
QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
|||||
Db 266 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 325
|||||
QY 121 AGARIHTNSWGAAYNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNCGGTISAPCTAKNAI 180
|||||
Db 326 AGARIHTNSWGAAYNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNCGGTISAPCTAKNAI 385
|||||
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFILSARSSLAPDSSF 240
|||||
Db 386 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFILSARSSLAPDSSF 445
|||||
QY 241 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLY 300
|||||
Db 446 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADVGUGY 505
|||||
QY 301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYFTATAGKPLKISLVNSDAPASTTA 360
|||||
Db 506 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYFTATAGKPLKISLVNSDAPASTTA 565
|||||
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNWDGRNNVNFVINAPOSQGTYYTIEVOAYN 420
|||||
Db 566 SVTLVNDLVLITAPNGTRYVGNDFTSAPFNNWDGRNNVNFVINSPOSQGTYYTIEVOAYN 625
|||||
QY 421 VPVGPQTFSLAIVN 434
|||||
Db 626 VPVGPQNFSLAIVN 639
|||||

RESULT 6
AAY17087
ID AAY17087 standard; protein: 639 AA.
XX AAY17087;
XX
XX 20-MAR-2003 (updated)
DT 21-JUL-1999 (first entry)
XX
XX An alkaline protease sequence from Bacillus species.
DE
XX Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
KW washing composition; oxidising agent.
XX
XX Bacillus sp.
XX
XX Key Location/Qualifiers
FH Misc-difference 1..639 /note= "all residues indicated as xaa are arbitrary
FT amino acids"
FT
FT
XX WO9918218-A1.
XX
XX 15-APR-1999.
XX
XX 07-OCT-1998; 98WO-JP04528.
XX
XX 07-OCT-1997; 97JP-0274570.
XX
XX (KAOS) KAO CORP.
XX
XX Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
PI Saeki K, Shikata S, Takaiwa M;
XX
XX WPI; 1999-287736/27.
XX
XX N-PSDB; AAX37277.
XX
XX Alkali protease from Bacillus used in washing powders
XX

PS Claim 3; Page 47-50; 71pp; Japanese.

CC The invention relates to alkaline proteases produced by strains of

CC Bacillus. The proteases ability to digest casein is not inhibited by

CC oleic acid and they have a high stability to oxidising agents. The

CC alkaline protease of the invention has the following properties: (a) it

CC is active over the pH range 4-13 and has at least 80% of its optimum

CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is

CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)

CC its ability to digest casein is not inhibited by oleic acid; (e) it has

CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be

CC used as enzymes in washing compositions for use in automatic dishwashers

CC and for washing clothes. The stability to oxidising agents allows the

CC enzyme to be an effective component of washing compositions including

CC bleaches. The present sequence represents an alkaline protease of the

CC invention.

CC (Updated on 20-MAR-2003 to correct DR field.)

XX Sequence 639 AA;

SQ Query Match 95.9%; Score 2155; DB 20; Length 639;

Best Local Similarity 96.3%; Pred. No. 2.2e-146;

Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60

DB 206 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 265

QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLOTLSQAYS 120

DB 266 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLOTLSQAYS 325

QY 121 AGARIHNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180

DB 326 AGARIHNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 385

QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGFIISARSLAPDSF 240

DB 386 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGFIISARSLAPDSF 445

QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALAGAADIGLY 300

DB 446 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALAGAADIGLY 505

QY 301 PNGNOGWRVTLDKSLNAVYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360

DB 506 PNGNOGWRVTLDKSLNAVYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 565

QY 361 SVTLVNDLVLITAPNGTQYVGNDFTPYNDNDGRRNNVENVFINAPQSGTYTIEVQAYN 420

DB 566 SVTLVNDLVLITAPNGTQYVGNDFTPYNDNDGRRNNVENVFINAPQSGTYTIEVQAYN 625

QY 421 VPVGQPTFFSLAIVN 434

DB 626 VPVGQPTFFSLAIVN 639

RESULT 7

AAV17088

ID AAY17088 standard; protein; 640 AA.

XX AC AAY17088;

XX DT 20-MAR-2003 (updated)

XX DT 21-JUL-1999 (first entry)

DE An alkaline protease sequence from Bacillus species.

KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;

KW washing composition; oxidising agent.

XX OS Bacillus sp.

FH Key Location/Qualifiers

FT Misc-difference 1..640

FT /note= "all residues indicated as Xaa are arbitrary

XX amino acids"

PN WO9918218-A1.

XX 15-APR-1999.

XX 07-OCT-1998; 98WO-JP04528.

XX 07-OCT-1997; 97JP-0274570.

XX (KAOS) KAO CORP.

PI Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;

PI Saeki K, Shikata S, Takaiwa M;

XX WPI: 1999-287736/27.

XX N-PSDB; AAX37278.

PT Alkali protease from Bacillus used in washing powders

XX Claim 3; Page 50-53; 71pp; Japanese.

PS The invention relates to alkaline proteases produced by strains of

CC Bacillus. The proteases ability to digest casein is not inhibited by

CC oleic acid and they have a high stability to oxidising agents. The

CC alkaline protease of the invention has the following properties: (a) it

CC is active over the pH range 4-13 and has at least 80% of its optimum

CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is

CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)

CC its ability to digest casein is not inhibited by oleic acid; (e) it has

CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be

CC used as enzymes in washing compositions for use in automatic dishwashers

CC and for washing clothes. The stability to oxidising agents allows the

CC enzyme to be an effective component of washing compositions including

CC bleaches. The present sequence represents an alkaline protease of the

CC invention.

CC (Updated on 20-MAR-2003 to correct DR field.)

XX Sequence 640 AA;

Query Match 95.9%; Score 2155; DB 20; Length 640;

Best Local Similarity 96.3%; Pred. No. 2.2e-146;

Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60

DB 207 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266

QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLOTLSQAYS 120

DB 267 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGSLPSNLOTLSQAYS 326

QY 121 AGARIHNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180

DB 327 AGARIHNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 386

QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGFIISARSLAPDSF 240

DB 387 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGFIISARSLAPDSF 446

QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALAGAADIGLY 300

DB 447 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALAGAADIGLY 506

QY 301 PNGNOGWRVTLDKSLNAVYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360

DB 507 PNGNOGWRVTLDKSLNAVYVNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 566

QY 361 SVTLVNDLVLITAPNGTQYVGNDFTPYNDNDGRRNNVENVFINAPQSGTYTIEVQAYN 420

QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQGTFFILSARSLAPDSSF 240
 Db 180 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQGTFFILSARSLAPDSSF 239
 QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLGY 300
 Db 240 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADVGLGF 299
 QY 301 PNGNQGWRVTLDKSLNVAVYNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 Db 300 PNGNQGWRVTLDKSLNVAVYNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 359
 QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVFNAPQSGTYTTEVQAYN 420
 Db 360 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVFNAPQSGTYTTEVQAYN 419
 QY 421 VPVGPOTFSLAIVN 434
 Db 420 VPVGPOTFSLAIVH 433

RESULT 10
 AAM89547

ID AAM89547 standard; Protein; 641 AA.

AC AAM89547;

DT 12-APR-1999 (first entry)

DE Bacillus Jp170 protease.

KW Protease; detergent; surfactant; leather processing; debittering;
 flavour.

OS Bacillus sp.

Key Location/Qualifiers
 1..33

FT Peptide /note= "signal peptide"

FT Region 34..208

FT Protein /note= "prepro region"

FT /note= "mature protein"

PN WO9856927-A2.

PD 17-DEC-1998.

PF 09-JUN-1998; 98WO-US12005.

PR 12-JUN-1997; 97US-0873479.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PI Christianson L, Sloma A;

XX WP1; 1999-080908/07.

XX N-PSDB; AAV82382.

PT Novel protease from *Bacillus subtilis* 1C20 - useful in laundry and
 dishwashing detergents and for leather processing

PS Claim 7; Page 53-54; 77pp; English.

CC This is the amino acid sequence of a novel protease of *Bacillus* sp.
 CC isolated gene (see AAV82382). The entire protein, including the
 CC signal peptide and prepro region, has 77% identity to alkaline
 CC protease Y (see AAM89548) from *Bacillus*. The invention provides
 CC vectors, recombinant host cells and methods for the recombinant
 CC production of the protease. The protease is used in laundry and
 CC dishwashing detergents, for institutional and industrial cleaning,
 CC and for leather processing, as well as for debittering and

CC enhancing the degree of hydrolysis of protein hydrolysates, for
 CC flavour development through hydrolysis of proteins, degradation of
 CC undesired peptides and in enzymatic synthesis of peptides. It has
 CC enhanced stability towards oxidation under alkaline conditions,
 CC e.g. towards bleaching agents of the peroxy type. The invention
 CC also provides mutant cells in which the protease activity is
 CC diminished. Such cells can be used for the production of
 CC heterologous recombinant proteins.

XX Sequence 641 AA;

Query Match

Best Local Similarity 94.6%; Score 2125.5; Db 20; Length 641;

Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGOGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db 209 NDVARGIVKADVAQNNFGLYGOGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 268
 QY 61 NANDTNGHGHVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLTFLFSQAYS 120
 Db 269 NANDPNGHGHVAGSVLGN-ATNKGMAQANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 327
 QY 121 AGARIHTNSWGAAYNGAYTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db 328 AGARIHTNSWGAAYNGAYTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 387
 QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQGTFFILSARSLAPDSSF 240
 Db 388 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAQGTFFILSARSLAPDSSF 447
 QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLGY 300
 Db 448 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADVGLGF 507
 QY 301 PNGNQGWRVTLDKSLNVAVYNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 360
 Db 508 PNGNQGWRVTLDKSLNVAVYNESSLSSTOKATYSFTATAGKPLKISLVWSDAPASTTA 567
 QY 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVFNAPQSGTYTTEVQAYN 420
 Db 568 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVFNAPQSGTYTTEVQAYN 627
 QY 421 VPVGPOTFSLAIVN 434
 Db 628 VPVGPOTFSLAIVH 641

RESULT 11
 AAM50090

ID AAM50090 standard; protein; 434 AA.

AC AAM50090;

DT 12-AUG-2002 (first entry)

DE Bacillus sp KSM-KP43 alkaline protease protein variant.

KW Alkaline protease; detergent; laundry; bleaching; dishwasher; mutant;
 mutin.

OS Bacillus sp.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 46

FT /label= Y,w,a,d,e,t,v,i,i,h,s,k,q,m,c
 /note= "as claimed in Claim 3"

FT Misc-difference 54

FT /label= Y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
 /note= "as claimed in Claim 3"

FT Misc-difference 57

FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a

FT FT /note= "as claimed in Claim 3"
 FT FT 66
 FT FT /label= e,d,s,q,a,t,l,m,c,v,g,i
 FT FT /note= "as claimed in Claim 3"
 FT FT 84
 FT FT /label= OTHER, R
 FT FT /note= "OTHER= deleted residue. Specifically described
 in Claim 1"
 FT FT 101..106
 FT FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
 FT FT /note= "as claimed in Claim 3"
 FT FT 104
 FT FT /label= OTHER, P
 FT FT /note= "OTHER= deleted residue. Specifically described
 in Claim 1"
 FT FT 107
 FT FT /label= k,r,a,s
 FT FT /note= "as claimed in Claim 3"
 FT FT 119
 FT FT /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 124
 FT FT /label= a,k
 FT FT /note= "as claimed in Claim 3"
 FT FT 136
 FT FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
 FT FT /note= "as claimed in Claim 3"
 FT FT 138
 FT FT /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 148
 FT FT /label= y,f,a,n,e,t,v,h,s,k,e,m,g,d,p,r,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 193
 FT FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
 FT FT /note= "as claimed in Claim 3"
 FT FT 195
 FT FT /label= y,f,a,h,e,t,v,h,s,k,e,m,g,d,p,r,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 205
 FT FT /label= y,w,a,d,e,t,v,l,i,h,s,k,q,m,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 247
 FT FT /label= w,f,a,r,e,t,v,l,i,h,s,e,m,c
 FT FT /note= "as claimed in Claim 3"
 FT FT 256
 FT FT /label= OTHER,a,s,e,v,l,r,e,d
 FT FT /note= "OTHER= deleted residue. Specifically described
 in Claim 1"
 FT FT 264
 FT FT /label= e,d,s,q,a,t,l,m,c,v,g,i
 FT FT /note= "as claimed in Claim 3"
 FT FT 257
 FT FT /label= v,i
 FT FT /note= "as claimed in Claim 3"
 FT FT 342
 FT FT /label= k,s,e,f,v,r,y,l,i,t,m,c,y,d,e,h,p,a
 FT FT /note= "as claimed in Claim 3"
 FT FT 369
 FT FT /label= OTHER, d
 FT FT /note= "OTHER= deleted residue. Specifically described
 in Claim 1"
 FT FT
 FT FT EP1209233-A2.
 FT FT
 FT FT 29-MAY-2002.
 FT FT
 FT FT 22-NOV-2001; 2001EP-0127851.
 FT FT
 FT FT 22-NOV-2000; 2000JP-0355166.
 FT FT 12-APR-2001; 2001JP-0114048.
 FT FT
 FT FT (KAOS) KAO CORP.

XX Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
 PI Okuda M, Saeki K;
 XX WPI; 2002-437518/47.
 XX
 XX New modified alkaline proteases useful in detergent compositions -
 PT Claim 1; Page -: 25pp; English.
 XX
 XX This invention describes novel Bacillus sp. alkaline proteases useful in
 CC detergent compositions, especially in laundry, bleaching or automatic
 CC dishwasher detergents. The novel proteases have an increased detergency
 CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
 CC sequence represents a variant of the alkaline protease Kp43 from
 CC Bacillus sp strain KSM-Kp43 created from the wild-type protease
 CC represented in AAM50080
 CC NOTE: This sequence is not represented in the specification but has
 CC been constructed from the sequence represented in record AAM50080.
 XX
 XX Sequence 434 AA;
 SQ
 Query Match 92.2%; Score 2071; DB 23; Length 434;
 Best Local Similarity 94.0%; Pred. No. 1.4e-140;
 Matches 408; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAFRGKITAIYALGRTN 60
 DB 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHEAXRGKITAIYALXRTN 60
 QY 61 NANDTNGHGHVAGSVGLNGSTNKGMAPOANLVFQSIMDSGGGLGSLNQLTFSQAYS 120
 DB 61 NANDTNGHGHVAGSVGLNGSTNKGMAPOANLVFQSIMDSGGGLGSLNQLTFSQAYS 120
 QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 DB 121 AGAXIHTNSWGAANGAYTTDSRNVDYVVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPFSGSYADNINHVAFSSRGPTKDGRIKPDVMAFGTIFILSARSLAPDSSF 240
 DB 181 TVGATENLRPFSEXSADNINHVAFSSRGPTKDGRIKPDVMAFGTIFILSARSLAPDSSF 240
 QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 DB 241 WANHDSXYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 QY 301 PNGNQGWGRVTLDKSLNVAYVNESSSLSTSKATYSFTATAGKPKLSILVWSDAPASTTA 360
 DB 301 PNGNQGWGRVTLDKSLNVAYVNESSSLSTSKATYSFTATAGKPKLSILVWSDAPASTTA 360
 QY 361 SVTLVNDLXLVITAPNGTOYVGNDFTPYNDNWDGRNNVNFVINAPOSCTVTIEVOAYN 420
 DB 361 SVTLVNDLXLVITAPNGTOYVGNDFTPYNDNWDGRNNVNFVINAPOSCTVTIEVOAYN 420
 QY 421 VPVGQPTFSLAIVN 434
 DB 421 VPVGQPTFSLAIVN 434
 RESULT 12
 AAM50084
 ID AAM50084 standard; protein; 433 AA.
 XX
 XX AAM50084;
 XX
 XX 12-AUG-2002 (first entry)
 DT Bacillus sp SD-521 (FERM BP-11162) alkaline protease protein fragment.
 DE Alkaline protease; detergent; laundry; bleaching; dishwasher.
 KW Bacillus sp.
 XX
 OS
 XX

```

PN  EP1209233-A2.
XX
PD  29-MAY-2002.
XX
PF  22-NOV-2001; 2001EP-0127851.
XX
PR  22-NOV-2000; 2000JP-0355166.
XX
PR  12-APR-2001; 2001JP-0114048.
XX
PA  (KAOS ) KAO CORP.
XX
PI  Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
PI  Okuda M, Saeki K;
XX
DR  WPI; 2002-437518/47.
XX
PT  New modified alkaline proteases useful in detergent compositions -
XX
PS  Claim 5; Page 16-18; 25pp; English.
XX
CC  This invention describes novel Bacillus sp. alkaline proteases useful in
CC  detergent compositions, especially in laundry, bleaching or automatic
CC  dishwasher detergents. The novel proteases have an increased detergency &
CC  (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC  sequence represents a fragment of the alkaline protease SD-521 from
CC  Bacillus sp strain SD-521 (FERM BP-11162) described in the method of the
CC  invention.
XX
SQ  Sequence 433 AA;

Query Match      88.9%; Score 1998.5; DB 23; Length 433;
Best Local Similarity 87.8%; Pred. No. 2.3e-135;
Matches 381; Conservative 28; Mismatches 24; Indels 1; Gaps 1;

QY  1 NDVARGIVKADVAQSSYGLYGOGQIVAVADTGLDTGRNDSSMHEAFRGKITAIYALGRTN 60
DB  1 NDVARGIVKADVAQNNYGLYGOGQVAVADTGLDTGRNDSSMHEAFRGKITAIYALGRTN 60

QY  61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB  61 NANDPNGHGTTHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 119

QY  121 AGARHTNSWGAANGAVTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
DB  120 AGARHTNSWGAAPVNGAYTANSRQVDEYVRNNDMTVLFAGNEGPNNGTISAPGTAKNAI 179

QY  181 TVGATENLRPSFGSYADNINHAQFSSRGPTKDGRIKPDVMAAGTIFILSARSLAPDSSF 240
DB  180 TVGATENYRPSFGLADPNHIAQFSSRGATRDGRIKPDVTAPGTIFILSARSLAPDSSF 239

QY  241 WANHSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB  240 WANYSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALITAGATDVLGY 299

QY  301 PNGNOGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPKLKSILVWSDAPASTTA 360
DB  300 PSGDQGWGRVTLDKSLNVAAYVNEATALTGOKATYSFTQAGKPKLKSILVWTDAPGSTTA 359

QY  361 SVTLVNDLDELVTAPNGQYVGNDFTPSYNDNWDGRNNVNFVINAPOSQGTYYTIEVOAYN 420
DB  360 SYTLVNDLDELVTAPNGQYVGNDFSYPDYNDNWDGRNNVNFVINAPOSQGTYYTIEVOAYN 419

QY  421 VPVGQPTFSLAIVN 434
DB  420 VPSGPQRFSLAIVH 433

RESULT 13
AAM50082
ID  AAM50082 standard; protein; 433 AA.
XX
AC
XX  AAM50082;
XX

```

```

DT  12-AUG-2002 (first entry)
XX
DE  Bacillus sp D6-(FERM P1592) alkaline protease protein fragment.
XX
KW  Alkaline protease; detergent; laundry; bleaching; dishwasher.
XX
OS  Bacillus sp.
XX
PN  EP1209233-A2.
XX
PD  29-MAY-2002.
XX
PF  22-NOV-2001; 2001EP-0127851.
XX
PR  22-NOV-2000; 2000JP-0355166.
XX
PR  12-APR-2001; 2001JP-0114048.
XX
PA  (KAOS ) KAO CORP.
XX
PI  Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;
PI  Okuda M, Saeki K;
XX
DR  WPI; 2002-437518/47.
XX
PT  New modified alkaline proteases useful in detergent compositions -
XX
PS  Claim 5; Page 13-15; 25pp; English.
XX
CC  This invention describes novel Bacillus sp. alkaline proteases useful in
CC  detergent compositions, especially in laundry, bleaching or automatic
CC  dishwasher detergents. The novel proteases have an increased detergency &
CC  (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This
CC  sequence represents a fragment of the alkaline protease E-1 from
CC  Bacillus sp strain D6-(FERM-P1592) described in the method of the
CC  invention.
XX
SQ  Sequence 433 AA;

Query Match      88.8%; Score 1994.5; DB 23; Length 433;
Best Local Similarity 87.6%; Pred. No. 4.5e-135;
Matches 380; Conservative 28; Mismatches 25; Indels 1; Gaps 1;

QY  1 NDVARGIVKADVAQSSYGLYGOGQIVAVADTGLDTGRNDSSMHEAFRGKITAIYALGRTN 60
DB  1 NDVARGIVKADVAQNNYGLYGOGQVAVADTGLDTGRNDSSMHEAFRGKITAIYALGRTN 60

QY  61 NANDTNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB  61 NANDPNGHGTTHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLFSQAWN 119

QY  121 AGARHTNSWGAANGAVTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
DB  120 AGARHTNSWGAAPVNGAYTANSRQVDEYVRNNDMTVLFAGNEGPNNGTISAPGTAKNAI 179

QY  181 TVGATENLRPSFGSYADNINHAQFSSRGPTKDGRIKPDVMAAGTIFILSARSLAPDSSF 240
DB  180 TVGATENYRPSFSGSIADPNHIAQFSSRGATRDGRIKPDVTAPGTIFILSARSLAPDSSF 239

QY  241 WANHSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
DB  240 WANYSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALITAGATDVLGY 299

QY  301 PNGNOGWRVTLDKSLNVAAYVNESSLSSTOKATYSFTATAGKPKLKSILVWSDAPASTTA 360
DB  300 PSGDQGWGRVTLDKSLNVAAYVNEATALTGOKATYSFTQAGKPKLKSILVWTDAPGSTTA 359

QY  361 SVTLVNDLDELVTAPNGQYVGNDFTPSYNDNWDGRNNVNFVINAPOSQGTYYTIEVOAYN 420
DB  360 SYTLVNDLDELVTAPNGQYVGNDFSYPDYNDNWDGRNNVNFVINAPOSQGTYYTIEVOAYN 419

QY  421 VPVGQPTFSLAIVN 434
DB  420 VPSGPQRFSLAIVH 433

```

RESULT 14

AAW89548
ID AAW89548 standard; Protein: 636 AA.

XX AAW89548;
XX
XX

DT 12-APR-1999 (first entry)
XX
XX

DE Bacillus sp. alkaline protease Y.
XX
XX

KW Alkaline protease Y; detergent; surfactant; leather processing;
KW debittering; flavour.
XX
XX

OS Bacillus sp.
XX
XX

PN W09856927-A2.
XX
XX

PD 17-DEC-1998.
XX
XX

PF 09-JUN-1998; 98WO-US12005.
XX
XX

PR 12-JUN-1997; 97US-0873479.
XX
XX

PA (NOVO) NOVO NORDISK BIOTECH INC.
XX
XX

PI Christianson L, Sloma A;
XX
XX

DR WPI; 1999-080908/07.
XX
XX

PT Novel protease from Bacillus subtilis LC20 - useful in laundry and
PT dishwashing detergents and for leather processing
XX
XX

PS Claim 3; Page 55-56; 77pp; English.
XX
XX

CC This is the amino acid sequence of a Bacillus sp. alkaline protease
CC Y that is said to have good alkali and surfactant resistance and
CC improved detergency. It shows 77% identity to a newly isolated
CC protease (see AAW89547) of Bacillus sp. JPI70 (NCIB 12513). The
CC invention provides vectors, recombinant host cells and methods for
CC the recombinant production of such proteases. The protease are
CC used in laundry and dishwashing detergents, for institutional and
CC industrial cleaning, and for leather processing, as well as for
CC debittering and enhancing the degree of hydrolysis of protein
CC hydrolysates, for flavour development through hydrolysis of
CC proteins, degradation of undesired peptides and in enzymatic
CC synthesis of peptides. They have enhanced stability towards
CC oxidation under alkaline conditions, e.g. towards bleaching agents
CC of the peroxy type. The invention also provides mutant cells in
CC which the protease activity is diminished. Such cells can be used
CC for the production of heterologous recombinant proteins.
XX
XX

SQ Sequence 636 AA;

Query Match 88.5%; Score 1989.5; DB 20; Length 636;
Best Local Similarity 87.3%; Pred. No. 1.7e-134;
Matches 379; Conservative 30; Mismatches 24; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60

DB 204 NDVARGIVKADVAQNNVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 263

QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120

DB 264 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAWN 322

QY 121 AGARIHTNSGAVNAGYATTDTSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180

DB 323 AGARIHTNSGAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNEGPNSGTISAPGTAKNAI 382

QY 181 TVGATENLRPSFGSIADNINHAQFSSRGPTKDCRIKPDVNAAGTFFILSARSLAPDSF 240

DB 181 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAPGTFFILSARSLAPDSF 442

QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300

DB 443 WANYSKYAYMGTSMTATPIVAGNVAQLREHFIRNRTGTPKPSLLKAALIAAGADVGLGY 502

QY 301 PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPKLKISILVNSDAPASTTA 360

DB 503 PSGDQGWRTVLDKSLNVAYVNEATATLGKATYSFOAQAGKPKLKISILVNTDAPGSTTA 562

QY 361 SVTLVNDLDTITAPNGTQYVGNDFTSYNDNDWGRNNVNFVINAPOSQGVTTIEVQAYN 420

DB 563 SVTLVNDLDTITAPNGQKYVGNDFSAPYDNDWGRNNVNFVINAPOSQGVTTIEVQAYN 622

QY 421 VPVGPOTFSLAIYN 434

DB 623 VPSGQRFSLAIYH 636

RESULT 15

AAW50083

ID AAW50083 standard; protein: 433 AA.

XX

AC AAW50083;

XX

DT 12-AUG-2002 (first entry)

XX

DE Bacillus sp Y-(FERM BP-1029) alkaline protease protein fragment.

XX

KW Alkaline protease; detergent; laundry; bleaching; dishwasher.

XX

OS Bacillus sp.

XX

PN EP1209233-A2.

XX

PD 29-MAY-2002.

XX

PF 22-NOV-2001; 2001EP-0127851.

XX

PR 22-NOV-2001; 2000JP-0355166.

XX

PR 12-APR-2001; 2001JP-0114048.

XX

PA (KAOS) KAO CORP.

XX

PI Hatada Y, Ogawa A, Kageyama Y, Sato T, Araki H, Sumitomo N;

XX

PI Okuda M, Saeki K;

XX

XX WPI; 2002-437518/47.

DR

XX New modified alkaline proteases useful in detergent compositions

XX

PS Claim 5; Page 15-16; 25pp; English.

XX

CC This invention describes novel Bacillus sp. alkaline proteases useful in

CC detergent compositions, especially in laundry, bleaching or automatic

CC dishwasher detergents. The novel proteases have an increased detergency

CC (34 - 38%) compared to prior art alkaline proteases (31 and 23%). This

CC sequence represents a fragment of the alkaline protease Ya from

CC Bacillus sp strain Y-(FERM BP-1029) described in the method of the

CC invention.

XX

SQ Sequence 433 AA;

Query Match 88.5%; Score 1987.5; DB 23; Length 433;

Best Local Similarity 87.3%; Pred. No. 1.4e-134;

Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60

DB 1 NDVARGIVKADVAQNNVGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60

QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIIMDSGGGLGGLPSNLQTLFSQAYS 120

DB 181 TVGATENLRPSFGSIADNINHAQFSSRGPTKDCRIKPDVNAAGTFFILSARSLAPDSF 240

DB 181 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAPGTFFILSARSLAPDSF 442

QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300

DB 443 WANYSKYAYMGTSMTATPIVAGNVAQLREHFIRNRTGTPKPSLLKAALIAAGADVGLGY 502

QY 301 PNGNOGWRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPKLKISILVNSDAPASTTA 360

DB 503 PSGDQGWRTVLDKSLNVAYVNEATATLGKATYSFOAQAGKPKLKISILVNTDAPGSTTA 562

Db 61 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSSGGLGLPSNLNTLFSQAWN 119
QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAAPVNGAYTANSRQVDEYVERNNDMTVLFAAGNEGPNSGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVYMAPGTFILSARSSSLAPDSSF 240
Db 180 TVGATENYRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVYMAPGTFILSARSSSLAPDSSF 239
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRCITPKPSLLKAALITAGADICLGY 300
Db 240 WANYNSKYAYMGTSMATPIVAGNVAQLREHFVKNRCITPKPSLLKAALITAGADICLGY 299
QY 301 PNGNOGWGRVTLDKSLNVAYVNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
Db 300 PNGDOGWGRVTLNKSINVAYVNEATATGQKATYSFOAQAGKPLKISLVWTDAPGSTTA 359
QY 361 SVTLVNDLVLITAPNGTOYVGNDFTSYNDNWDGNNVNFVFNAPQSGTYTIEVOAYN 420
Db 360 SVTLVNDLVLITAPNGOKYVGNDFSYNDNWDGNNVNFVFNAPQSGTYTIEVOAYN 419
QY 421 VPGPQTFSLAIVN 434
Db 420 VPSGPQTFSLAIVH 433

Search completed: July 25, 2003, 18:59:39
Job time : 43.0461 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 28, 2003, 01:10:39 ; Search time 375.933 Seconds
(without alignments)
3116.395 Million cell updates/sec

Title: US-09-985-689A-1
 Perfect score: 2247
 Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVOAYNPVPGPOTFSLAIVN 434

Scoring table:	BLOSUM62
	Xgapop 10.0, Xgapext 0.5
	Ygapop 10.0, Ygapext 0.5
	Fgapop 6.0, Fgapext 7.0
	Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

total number of hits satisfying	chosen parameters:	5105512
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1
20	1	1
21	1	1
22	1	1
23	1	1
24	1	1
25	1	1
26	1	1
27	1	1
28	1	1
29	1	1
30	1	1
31	1	1
32	1	1
33	1	1
34	1	1
35	1	1
36	1	1
37	1	1
38	1	1
39	1	1
40	1	1
41	1	1
42	1	1
43	1	1
44	1	1
45	1	1
46	1	1
47	1	1
48	1	1
49	1	1
50	1	1
51	1	1
52	1	1
53	1	1
54	1	1
55	1	1
56	1	1
57	1	1
58	1	1
59	1	1
60	1	1
61	1	1
62	1	1
63	1	1
64	1	1
65	1	1
66	1	1
67	1	1
68	1	1
69	1	1
70	1	1
71	1	1
72	1	1
73	1	1
74	1	1
75	1	1
76	1	1
77	1	1
78	1	1
79	1	1
80	1	1
81	1	1
82	1	1
83	1	1
84	1	1
85	1	1
86	1	1
87	1	1
88	1	1
89	1	1
90	1	1
91	1	1
92	1	1
93	1	1
94	1	1
95	1	1
96	1	1
97	1	1
98	1	1
99	1	1
100	1	1

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```

MODEL=frameat_p2n_model -DEV=xlp
Q=/cgn2_1/USF70_spool/US0985689/runat_25072003_143031_821/app_query.fasta_1.1166
DB=N_Geneseq_19Jun03 -QMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LoopCPL=0
LoopEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US0985689.@cgn_1_1_606_grunat_25072003_143031_821 -NCPU=6 -ICPU=3
NO_MMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

```

atabase : N Geneseq 19.Jun03.*

1:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA2002.DAT:*
25:	/SIDSL1/gcgdata1/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Query			DB	ID	Description
			Match	Length				
1	1	2247	100.0	1923	20	AAK37278	Bacillus alkaline	
2	2	2242	99.8	1923	20	AAK37277	Bacillus alkaline	
3	3	2183	97.2	1920	20	AAK37279	Bacillus alkaline	
4	4	2125.5	94.6	3003	20	AAW822382	Bacillus jiri70 pro	
5	5	1986.5	88.4	1299	13	AAQ27516	Thermococcus prote	
6	6	452.5	20.1	1977	18	AAW85667	Thermococcus prote	
7	7	452.5	20.1	1977	20	AAW05926	W09456929 Seq ID 1	
8	8	416.5	18.5	1236	20	AAW05920	Hyperthermostable	
9	9	416.5	18.5	1566	18	AAW85668	Pyrococcus furiosu	
10	10	416.5	18.5	1962	18	AAW85695	Pyrococcus furiosu	
11	11	416.5	18.5	1962	20	AAW05929	Hyperthermostable	
12	12	403.5	18.0	1977	18	AAW85669	Protease coding se	
13	13	372	16.6	2121	23	ABL54900	T. yonsei subtile	
14	14	346	15.4	2539	18	AAW61454	Streptomyces virid	
15	15	346	15.4	2809	18	AAW61455	DhpA-mel chimeric	
16	16	338	15.0	135638	25	ABX34289	S. atroolivaceus 1	
17	17	332.5	14.8	1329	24	ABK74643	Bacillus lichenifo	
18	18	310.5	13.8	2532	13	AAQ29134	Encodes RP-III res	
19	19	307.5	13.7	1859	18	AAW85677	Thermococcus prote	
20	20	307	13.7	2835	17	AAW08141	Hyperthermostable	
21	21	307	13.7	4765	17	AAW08132	Protease gene. Pyr	
22	22	307	13.7	4765	18	AAW85670	Pyrococcus furiosu	
23	23	307	13.7	4765	20	AAW05921	W09456926 Seq ID 5	
24	24	306.5	13.6	1560	25	AAW375569	Streptomyces virid	
25	25	306.5	13.6	59816	25	ABZ37515	Streptomyces virid	
26	26	306.5	13.6	59816	25	ABZ37516	Streptomyces virid	
27	27	301.5	13.4	3413	20	AAW72330	F. balustinum Cp70	
28	28	296	13.2	898	17	AAW08131	Hyperthermostable	
29	29	290.5	12.9	564	17	AAW08134	DNA sequence. Pyr	
30	30	290.5	12.9	564	18	AAW85676	Thermococcus prote	
31	31	276	12.3	1306	24	ABL55784	Bacillus sp. TV145	
32	32	276	12.3	1330	24	ABL55787	Bacillus sp. prote	
33	33	275	12.2	546	23	ABL53453	T. yonsei subtile	
34	34	265.5	11.8	2273	11	AAW04339	Aquaricine I gene.	
35	35	265.5	11.8	2274	12	AAQ25838	Aquaricin I coding	
36	36	265.5	11.8	2274	16	AAQ25839	Aquaricin I gene.	
37	37	263	11.7	4740	25	ABZ58957	Group B Streptococ	
38	38	256.5	11.4	10216	17	AAW39279	Transposon deliver	
39	39	256	11.4	1485	24	ABK74647	Bacillus lichenifo	
40	40	256	11.4	4650	24	ABW71526	Streptococcus poly	
41	41	256	11.4	4710	24	ABW71162	Streptococcus poly	
42	42	256	11.4	4734	24	ABW70525	Streptococcus poly	
43	43	256	11.4	4770	24	ABW69191	Streptococcus poly	
44	44	256	11.4	215561	24	AAW71527	Streptococcus poly	
45	45	253.5	11.3	3117	20	ABW71268	Streptococcus poly	

ALIGNMENTS

AAK37278 standard; DNA; 1923 BP.

AAK37278;

20-MAR-2003 (updated)

21-JUL-1999 (first entry)

Bacillus alkaline protease encoding DNA.

Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
washing composition; oxidising agent; ss.

Bacillus sp.

W03918218-A1

XX PD 15-APR-1999.

XX PF 07-OCT-1998; 98WO-JP04528.

XX PR 07-OCT-1997; 97JP-0274570.

XX PA (KAOS) KAO CORP.

XX PI Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;

XX PI Saeiki K, Shikata S, Takaiwa M;

XX DR WPI; 1999-287736/27.

XX DR P-PSDB; AAY17088, AAY17090.

XX PT Alkali protease from Bacillus used in washing powders

XX PS Disclosure; Page 58-63; 7lpp; Japanese.

CC The invention relates to alkaline proteases produced by strains of

CC Bacillus. The proteases ability to digest casein is not inhibited by

CC oleic acid and they have a high stability to oxidizing agents. The

CC alkaline protease of the invention has the following properties: (a) it

CC is active over the pH range 4-13 and has at least 80% of its optimum

CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is

CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)

CC its ability to digest casein is not inhibited by oleic acid; (e) it has

CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be

CC used as enzymes in washing compositions for use in automatic dishwashers

CC and for washing clothes. The stability to oxidizing agents allows the

CC enzyme to be an effective component of washing compositions including

CC bleaches. The present sequence represents an alkaline protease encoding

CC DNA.

CC (Updated on 20-MAR-2003 to correct DR field.)

XX SQ Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 other;

Alignment Scores:

Pred. No.: 4,2e-164 Length: 1923

Score: 2247.00 Matches: 434

Percent Similarity: 100.00% Conservatives: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 20 Gaps: 0

US-09-985-689A-1 (1-434) x AAX37278 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20

DB 619 AATGATGTTGGCGGTGAATTTGTCAAGCGGATGTGGCTCAGGACGCTACGGGTTGTAT 678

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

DB 679 GGACAGGACAGATCGTAGCGGTTCGCCATACAGGCTTGATACAGTCCCAATGACAGT 738

QY 41 SerMethisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

DB 739 TCGATGATGAAGCTTCGCGGGGAAATTTACTGCATTATATGCAATGGGACGACGAAT 798

QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80

DB 799 AATGCCAATGATACGAATGGTCATGGTACGCGATGTGGCTTCGGCTTCCTATTAGGAACGGC 858

QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100

DB 859 TCCACTAATAAGGAATGGCGCTCAGGGGATCTAGTCTTCCAAATCTATCATGATAGC 918

QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120

DB 919 GGTGGGGACATTGGAGGACTACCTTCGAATCTGCAACCTTATTCAGCCCAAGCATACAGT 978

QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140

DB 979 GCTGGTGCCAGAAATTCATACAACTCTCTGGGGAGCAGCAGTGAATGGGCTTACACAACA 1038

QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160

DB 1039 GATTTCAGAAATGTGATGATGTGCGCAAAATGATATGACGATCTTTTCGCTGCC 1098

QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180

DB 1099 GGAATGAAGGACCGAAGCGGGAACCATCACTGCACGAGCAGCAGTAAAAATGCAATA 1158

QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

DB 1159 ACAGTCGGAGCTACGGGAAACCTCCGCCAAAGCTTTGGTCTCTATGCGGACAAATATCAAC 1218

QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220

DB 1219 CATGTGGCAGATTCTCTTCACGTGGCCGCAAAAGGATGGACGATCAACCGGATGTC 1278

QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240

DB 1279 ATGCACCGGGAAGCTTCATCACTATCAGCAAGATCTCTCTTGCACCGGATTCCTCTTC 1338

QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260

DB 1339 TGGCGGAACCATGACAGTAAATATGCATATGCGGTGGAACTGCTCCATGCTACCCGATC 1398

QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280

DB 1399 GTTGCTGGAAACGTGGCACAGCTTCGTGAGCATTTGTGAAAAACAGAGGCATCACACCA 1458

QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300

DB 1459 AAGCTTCTCTATTAAAAAGCGCACTGATTGGCGGTGCACTGACATCGGCGCTGGGTAC 1518

QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320

DB 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATGGATAAATCCCTGAACGTTGCCTAT 1578

QY 321 ValAsnGluSerSerSerLeuSerThrGlnLysAlaThrTyrSerPheThrAlaThr 340

DB 1579 GTGAACGAGTCCAGTTCTATCCACCGCAAAAGCGACGCTACTCGTTTACTGCTACT 1638

QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360

DB 1639 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGTGATGCCCTGCCGACACAACTGCT 1698

QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380

DB 1699 TCCGTAAACGCTGTCAATGATCTGGACCTTGTCATTACCGCTCCAAATGGCACACGAT 1758

QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400

DB 1759 GTAGGAATGACTTTACTTCGCATACAATGATGATGATGATGATGATGATGATGATGAT 1818

QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420

DB 1819 AATGTATTATTATTCACACCAACAGCGGACGATACATTAAGGTACAGCTTATAAC 1878

QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434

DB 1879 GTACCGGTGGACACAGACCTTCTCGTTGGCAATTTGTGAAT 1920

RESULT 2

AAX37278

ID AAX37279 standard; DNA; 1923 BP.

XX AC AAX37279;

XX XX 20-MAR-2003 (updated)

DT 21-JUL-1999 (first entry)

XX DE Bacillus alkaline protease encoding DNA.

XX XX

KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX Bacillus sp.
 XX WO9918218-A1.
 XX 15-APR-1999.
 XX 07-OCT-1998; 98WO-JP04528.
 XX 07-OCT-1997; 97JP-0274570.
 XX (KAOS) KAO CORP.
 XX Hitomi J, Kageyama Y, Kubota H, Nomura M, Okuda M;
 PI Saeki K, Shikata S, Takaiwa M;
 XX WPI: 1999-287736/27.
 DR P-PSDB; AA17091.
 XX Alkali protease from Bacillus used in washing powders
 PT Disclosure; Page 63-68; 71pp; Japanese.
 XX The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA.
 CC (Updated on 20-MAR-2003 to correct DR field.)
 XX

SO Sequence 1923 BP: 578 A; 417 C; 474 G; 454 T; 0 other;

Alignment Scores:
 Pred. No.: 1,02e-163 Length: 1923
 Score: 2242.00 Matches: 433
 Percent Similarity: 99.77% Conservative: 0
 Best Local Similarity: 99.77% Mismatches: 1
 Query Match: 99.78% Indels: 0
 DB: 20 Gaps: 0

US-09-985-689A-1 (1-434) x AAX37279 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 DB 619 AATGATGTTGGCGTGGAAATGTCAAGCGGATGGTCAAGCGAGCTACGGGTGTAT 678
 QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 679 GGACAAGGACAGATCGGTAGCGGGTTGCCGATACAGGGCTTGATACAGGTGCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 739 TCGATGATGAAGCTTCCCGGGGAAATTTACTGCAATATATGCAATGGGACGGAGAT 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 799 AATGCCAATGATACGAATGTTGATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 858
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 859 TCACACTAATAGAGATGGCGCCCTCAGCGCAATCTAGTCTTCCAATCTATCATGATAGC 918

QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 DB 919 GGTGGGGGACTTGGAGGACTACTCTCGAATCTCAAACCTTATTACGCCAAGCATACAGT 978
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
 DB 979 GCTGGTGCAGAAATTCATAAACTCTCTGGGGAGCAGCATGAATGGGGCTTACACAACA 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTCGCTGCC 1098
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 DB 1099 GGAATGAAGGACCGAAGCGGACCATCATGTCACCGACGACAGTAAATAATGCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 DB 1159 ACAGTCGAGCTACGGAACCTCCGCCCAAGCTTTGGGTCTTATCGGACAAATATCAAC 1218
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 DB 1219 CATGTGGCACAGTCTCTTCACGTGACCGCAAGGATGGACGATCAAAACCGGATGTC 1278
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 DB 1279 ATGGCACCGGAACGTTTCATCTACGCAAGATCTTCTTTCGACCGGATTCCTCCTTC 1338
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 DB 1339 TGGCGCAACCATGACAGTAAATATGCATACATGSGTGGGAACGTCATGCTACACCGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 DB 1399 GTTCTGGAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAAACAGAGGCGATCACACCA 1458
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 DB 1459 AAGCTTCTCTATTAAAAGCGGCACTGATTGCCGGTGCAGCTGACATCGGCTTGGCTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 DB 1519 CCGAACCGTAAACCAAGGATGGGACGAGTGACATTTGGATAAATCCCTGAACCTTGCCTAT 1578
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 DB 1579 GTGACGAGTCCAGTTCCTCTATCCACAGCCCAAAAGCGAGCTACTCGTTTACTGCTACT 1638
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 DB 1639 GCGGGCAAGCCTTTGAAAATCTCCCTGCTGATGGTCTGATGCCCTCGGACGACAACTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 DB 1699 TCCGTACAGCTTGTCAATGATCTGGACCTTGTCTATTACCGCTCCAAATGGCACACAGTAT 1758
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 DB 1759 GTAGGAAATGACTTTACTTCGCCATCAATGATAACTGGGATGGCGCCCAATACGTAGAA 1818
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValIleGlnAlaTyrAsn 420
 DB 1819 AATGTATTTATTATGCACCAAGCGGACGCTATACAAATTTGAAGTACAGGCTTATAAC 1878
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 1879 GTACCGGTTGGACCAAGCAACTCTCTCGTTGGCAATTTGTGAAT 1920

RESULT 3
 AAX37277
 ID AAX37277 standard; DNA: 1920 BP.
 XX
 AC AAX37277;

RESULT 4
AAV82382
ID AAV82382 standard; DNA; 3003 BP.
XX AC AAV82382;
XX 12-APR-1999 (first entry)
XX DE Bacillus Jp170 protease gene.
XX KW Protease; detergent; surfactant; leather processing; debittering;
XX KW flavour; ss.
XX OS Bacillus sp.
XX FH Key Location/Qualifiers
XX CDS 846..2771
XX FT /*tag= a
XX FT sig_peptide 846..944
XX FT /*tag= b
XX FT mat_peptide 1470..2768
XX FT /*tag= c
XX PN W09856927-A2.
XX 17-DEC-1998.
XX 09-JUN-1998; 98WO-US12005.
XX 12-JUN-1997; 97US-0873479.
XX (NOVO) NOVO NORDISK BIOTECH INC.
XX Christianson L, Sloma A;
XX WPI: 1999-080908/07.
XX P-PSDB: AAV89547.
XX Novel protease from Bacillus subtilis LC20 - useful in laundry and
XX dishwashing detergents and for leather processing
XX Claim 11; Page 52-53; 77pp; English.
XX This nucleotide sequence encodes a novel protease (see AAV82382) of
XX Bacillus sp. Jp170 (NCIB 12513). The sequence in plasmid p170BAN
XX is contained in Bacillus subtilis LC20 NRRL B-21680. The protease
XX gene was isolated from chromosomal DNA of Jp170 following
XX preparation of probes based on protease N-terminal and internal
XX peptides (see AAV89549-50), screening of chromosomal libraries,
XX isolation of the 3' end of the gene by inverse PCR (see AAV82410-11),
XX reconstruction of 5' and 3' ends and PCR amplification (see
XX AAV82412-16). Claimed recombinant host cells can be used in a method
XX for producing the protease. The protease is used in laundry and
XX dishwashing detergents, for institutional and industrial cleaning,
XX and for leather processing, as well as for debittering and
XX enhancing the degree of hydrolysis of protein hydrolysates, for
XX flavour development through hydrolysis of proteins, degradation of
XX undesired peptides and in enzymatic synthesis of peptides. It has
XX enhanced stability towards oxidation under alkaline conditions. e.g.
XX towards bleaching agents of the peroxy type. The invention
XX also provides mutant cells in which the protease activity is
XX diminished. Such cells can be used for the production of
XX heterologous recombinant proteins.
XX SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 other;
Alignment Scores:
Pred. No.: 1.68e-154 Length: 3003
Score: 2125.50 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.55% Mismatches: 8
Query Match: 94.59% Indels: 1

DB:	20	Gaps:	1
US-09-985-689A-1 (1-434) x AAV82382 (1-3003);			
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20			
DB 1470 AATGACGTGGCCCGTGGCATTTGTAAAGCAGACGTCGCAAAATAAATTTGGCTTATAT 1529			
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40			
DB 1530 GGACAAGCAGACAGATTGTAGCAGTTCTGATCTGGCTTGATACAGCAAGAAATGACAGT 1589			
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60			
DB 1590 TCGATGTCATGAAGCATTCGCGGTAAAGATTACCGCACTATATGCACTGGCGCAACCAAT 1649			
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80			
DB 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGGATCTGTGTAGAAAT--- 1706			
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100			
DB 1707 GCTACAAATAAAGGATGGCACCACCAATCTAGTCTTCAATCTATTATGGATAGT 1766			
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120			
DB 1767 GGTGGAGGCTGGGAGGACTACCTGCTAATCTACAAACATTATTTCAGTCAAGCATATAGT 1826			
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAlaGlyValAlaTyrThr 140			
DB 1827 GCTGGAGCGAGAAATTCATACGAATTCATGGGGGCTCCAGTAAACGGTGCCTATACGACA 1886			
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160			
DB 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAATGATATGACATTCTTTTCGGGCC 1946			
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180			
DB 1947 GGAATGAGGACGAGGTAGCGGTACCAATCAGTCACCAAGCAAGCAAGCAAGCAAGT 2006			
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200			
DB 2007 ACAGTGGGGCAACCCGAAACCTACGTCCAAGCTTCGATCTTATTCGGGATATATTAAC 2066			
QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220			
DB 2067 CATGTTGCTCAATCTCTTCACGAGGTCCTACTAGAGATGGACGATTAAGCCGACGCTC 2126			
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240			
DB 2127 ATGGCACCAAGTACGTATATTCCTCTGCTAGATCATCATAGCTCCAGATTCCTCATTC 2186			
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260			
DB 2187 TGGGCAACCATGATAGTAAATATGCTACATGGTGGTACTTCTATGGCTACTCCAAT 2246			
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280			
DB 2247 GTAGCAGGTAATGTCACAAATTAAGGGACCATTTGTGAAAAAATAGAGGGGTAACTCCT 2306			
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300			
DB 2307 AAGCCTTCCTTTTAAAGCTGCTTTAATGTCAGGTGCTCCGGATGTTGGACTTGGCTTT 2366			
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320			
DB 2367 CCAATGTTACCAAGGATGGGAGAGATTAACGTTAGATAAAATCCCTAAATGTCGCAAT 2426			
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340			
DB 2427 GTGAATGAACAGCGCCCTTTATCAACAAGTCAAAAGCAACATATTTCGTTTACGGCTCAA 2486			
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360			

2487 GCTGGTAAACCCCTTAAATAATATCACTTCTTGGTCAGATGCCACAGGTAGCACGCGCA 2546
 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 2547 TCACCTACTTGTAGTGAATGATTAGACTTACTAGCTTACTGACCAAAATGGAATAATAC 2606
 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
 2607 GTCCGAATGACTTTACAGCACCGCTATGATACAAATGGGATGGCAGAAACACGTTGGAA 2666
 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrIleGluValGlnAlaTyrAsn 420
 2667 AATGTGTTTATCAATGCTCTCAAGCGAAGCTATACAGTGAAGTGCAGGCTTACANT 2726
 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 2727 GTACCACTAAGTCGCAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 5

AAQ27516 standard; DNA; 1299 BP.

XX AAQ27516;

XX 05-FEB-1993 (first entry)

XX Alkali-protease Ya enzyme gene.

XX Alkali resistance; surface active agent resistance;
 KW detergent improver; ss.

XX Bacillus sp. Y.

XX Key Location/Qualifiers

XX CDS 1..1299

XX /*tag= a

XX JP04197182-A.

XX 16-JUL-1992.

XX 28-NOV-1990; 90JP-0327110.

XX 28-NOV-1990; 90JP-0327110.

XX (LLOY) LION CORP.

XX WPI; 1992-288440/35.

XX P-PSDB; AAR26274.

XX DNA coding alkali-protease Ya enzyme - has good alkali and
 PT surfactant resistance and improves detergency

XX Claim 3; Page 2; 17pp; Japanese.

XX The sequence is that of the alkali-protease Ya enzyme gene which can
 CC be used in the recombinant production of Ya enzyme. Ya enzyme is
 CC excellent in alkali resistance and surface active agent resistance
 CC and improves detergency.

XX Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 other;

Alignment Scores:

Pred. No.: 3,27e-144 Length: 1299
 Score: 1986.50 Matches: 379
 Percent Similarity: 94.01% Conservative: 29
 Best Local Similarity: 87.33% Mismatches: 25
 Query Match: 88.41% Indels: 1
 DB: 13 Gaps: 1

US-09-985-689A-1 (1-434) x AAQ27516 (1-1299)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 DB 1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTGACAAACAATACGATTATAT 60
 QY 21 GlyGlnGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 61 GCACAAGGTCAACTAGTTGCAGTAGCGGACACAGGCTTAGATACAGGTCCTAAGCATAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 121 TCTATGCTAAGCAATTCGCGGAAATACACGCTCTTTACCGGTAGGAGAACTAAT 180
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 181 AATCGAGTGTCCGAATGGCATGGCACACATGTAGCAGGTCTCTGTTACTGTGTAAT 237
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAAGTCTGCTTCAATCTATTATGATAGC 297
 QY 101 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 DB 298 AGCGAGGATTAGTGGCTTACCATCGAACTTAATACGTTATTAGTCAAGCTTGAAT 357
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
 DB 358 GCTGGAGCAAGAATTCATCTACTCTTGGGGAGCCCCAGCAATAATGAGCGTACACTGCT 417
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 418 AACTCGAGCAAGTGGATGATGATGTTTCGAAATAATGATATGACGGTACTTTTGCAGCT 477
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 DB 478 GGTAATGAAGTCTTAATTCAGAGACAATAGTCTCCAGGTACAGCGCAAAATGCTATT 537
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 DB 538 ACGGTGCGCGCAACGGAAACTATCGCCCAAGCTTCGGTTCGATAGCAGATAACCCAAAT 597
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 DB 598 CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCTGACGTA 657
 QY 221 MetAlaProGlyThrPheIleLeuSerSerAlaArgSerSerLeuAlaProAspSerPhe 240
 DB 658 ACAGCTCTCTGGAACTTTATTTATCAGCAGCTTCTTCCCTAGCTCCAGACTCTTCGTT 717
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 DB 718 TGGCGCAATTAACAGATAAATACGCGTATATGGCGGTACCTCCAGCAGACCTATT 777
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 DB 778 GTTGCAGGAATTCGCGCAATTACGTGAGCATTTTATAAAAAATAGAGGTATTACTCT 837
 QY 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 DB 838 AAGCCTCTTTAATAAAGCTGCACCTATCCCTGGTGTACTGATGTTGTTTAGGATAT 897
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 DB 898 CCTAGTGTGTACCAAGGCTGGGGGCGTGTACTCTAGATAAATCTTAATGTAGCGTAT 957
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 DB 958 GTCAATGAAGCAACTGCTATAGCCACGACGACGACGACGACGACGATATTCGTCCAGCACAA 1017
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 DB 1018 GCGGGTAAACCTTTAAAAATCTCGTTAGTATGGACAGATGCTCTCTGGAAGTACACACTGCA 1077
 QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380

Db	1312	GACGACTACGCCAAGCTCACCTTCACCGGCTCGTCCGACAGGAAGCGCACCCAC	1371
Qy	336	SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTTPSerAspAlaPro	355
Db	1372	ACCTTCGACGTCAGCGCGCCACCTTCGTGACCGCCACCCCTCTACTGGAC	1422
Qy	356	AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro	375
Db	1423	-----ACGGGCTCGAGCGACATCGACCTCTACCTCTACGACCC	1461
Qy	376	AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTTPaspGly	395
Db	1462	AACGGGAACAG--GTTGACTACTCTACACCGCCTACTAC	1500
Qy	396	ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGlu	415
Db	1501	-----GGCTTCGAGAAGTTCGGCTACTACAACCCGACCGCGGACCTCGACGGTCAAG	1554
Qy	416	ValGlnAlaAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn	434
Db	1555	GTCTGACGTACAAAG-----GGCGGGCGAACTACCAAGTTCGACGTCGTCAAC	1602

Alignment Scores:

Pred. No.:	3_92e-23	Length:	1236
Score:	416.50	Matches:	141
Percent Similarity:	42.55%	Conservative:	59
Best Local Similarity:	30.00%	Mismatches:	148
Query Match:	18.54%	Indels:	122
DB:	20	Gaps:	20

US-09-985-689A-1 (1-434) x AAX05920 (1-1236)

QY	12	ValAlaGlnSerSerTyrGly-LeuTyr	-----GlyGlnGlyG1 24
DB	24	GTCTCAGCTCAAGTTATGCGCAACITACGTTTGGGAACCTGGGATATGATGGTTCTCGAAT	83
QY	24	nileValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44	
DB	84	CACAATAGCAATAATTGACACTTGAATTGAC	125
QY	44	uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn--	630
DB	126	AGATCTCCAGAAAGTA	170
QY	64	-----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78	
DB	171	TAGGAGTTATCATACGATGACCATGGACATCGAATCATGTAGCTTCAATACGACCTGG	230
QY	78	YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94	
DB	231	TACTGGACGACGAGTAATGCGAAGTACGAAGGAATGCTCCAGGAGCTAAGCTGCGGG	290
QY	94	eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuG1 112	
DB	291	AATTAAAGTTCTAGTGCCTGCTGTTCTGGAAGCATATCTACTATAAATAAGCGAGTTGA	350
QY	112	nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132	
DB	351	GTGGCCGTTGATACAAAGATAGTACGGAATTAAGGTCAFTAACTTCTCTGGTTC	410
QY	132	a-----AlaValAsnGlyAlaTyrThrTh 140	
DB	411	AAGCCAGAGCTCAGATGTAAGTCTAGTCAGGCTGTTAATGACGCTGGGATGC	470
QY	140	rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaA 160	
DB	471	T-----GGATAGTTGTTGGTTCGCGC	494
QY	160	aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180	
DB	495	TGGAACAGTGGACCTTAACAAGTATACAATCGGTTCTCCACGAGCTGCAACCAAGTTAT	554
QY	180	eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200	
DB	555	TACAGTTGGAGCC-----GTTGACAAGTATGA	581
QY	200	nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220	
DB	582	TGTTATAACAAGCTTCTCAAGCAGAGGCCCACTGCACAGCGGAGCTTAAGCCTGAGGT	641
QY	220	lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240	
DB	642	TGTTGTCGCCAGAACTGGATAATTGCTGTCGACGAGCAAGT-----GCAACTACGAT	692
QY	240	eTIPAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI 260	
DB	693	GGGTCAACCAATTAATCACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCTCA	752
QY	260	eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280	
DB	753	CGTAGCTGTTATGACAGCCCTCTTGTCTCAA-----GCACACCC	791
QY	280	oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla-----	293
DB	792	GAGCTGGACTCCAGACAAATAAAACAGCCCTCAATGAACCTCTGATATCGTAAGCC	851

PS Claim 7; Page 90-91; 159pp; Japanese.

XX This sequence represents the coding sequence for the protease from
CC *Pyrococcus furiosus* DSM-3638. This sequence encodes a protease of the
CC invention. The proteases of the invention have extremely high thermal
CC stability. The proteases can be used as research reagents, and
CC industrially in the food, drug and chemical industries.

XX Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 1 other;
SQ

Alignment Scores:
Pred. No.: 5,18e-23 Length: 1566
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 18 Gaps: 20

US-09-985-689A-1 (1-434) * AAT85668 (1-1566)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyCl 24
Db 24 GTCTGCAGCTCAAGTTATGGCAACTTACGTTGGAACTGGGATATGATGTTCTTGAAT 83
QY 24 nLeValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisL 44
Db 84 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAlaAsn-- 63
Db 126 AGATCTCCAAGGAAGAAGTA-----ATTGGTGGGTGATGTTTGTCAATG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuL 78
Db 171 TAGGAGTTATCCATACGATGACCATGACATGGAACATCATGTAGTCTTAATAGCAGCTGG 230
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAGGAATGCTCCAGGAGTAACTGAGGAGTTGA 290
QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuL 112
Db 291 AATTAAAGTTCTAGTCCGATGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 350
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGGCGGTGATAACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTGTTTC 410
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 411 AAGCCAGCTCAGATGGTACTGAGCTCTAAGTACAGCTGTTAATGACAGCTGGGATGC 470
QY 140 rAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAlaL 160
Db 471 T-----GGATAGTGTGTTGTTGTTGCGGC 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
Db 495 TGGAAACAGTGGACCTAACAGTATACATCCGTTCTCCAGCAGCTGCACAAAGTTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAAGTATGA 581
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATATAACAAGCTTCTCAAGCAGAGGGCCACTGCAGCGGCGGCTTAAGCTGAGGT 641
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240
Db 642 TGTGTCTCCAGGAACCTGGATAATTGCTCCAGAGCAAGT-----GGAACCTAGCAT 692
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI 260

QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 852 AGATGAATAGCCGATAGCTAGCGTGCA-----GGTAGGGT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 891 TAATGCATACAGGCTATAAAC-----TACGATAACTATGCAAAAGCTAGTGTTCACCTGG 944
QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGCAGCCAACTCAGCTCGTTATAGCGGAGCTTCGTCGT 1004
QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATATACTGGGACAATGCCAAT----- 1035
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1036 -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAACACAG-----GTTGACTACTCTTA 1091
QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCCCTACTAT-----GGATTCCAAAAGGTGTGTTATTA 1127
QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCAACTGATGGAACATGGAACAATTAAGGTGTGAAGCTACAG-----GGAAG 1178
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TCCAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 9
AAT85668
ID AAT85668 standard; DNA: 1566 BP.
AC AAT85668;
DT 20-APR-1998 (first entry)
XX *Pyrococcus furiosus* protease coding sequence.
DE Protease; research reagent; thermal stability: *pyrococcus furiosus*; ss.
XX *Pyrococcus furiosus* DSM-3638.
OS
FH Key Location/Qualifiers
FT CDS 1..1566
FT /*tag= a
FT /transl_except= (pos: 1282..1284, aa: Xaa)
FT /note= "Xaa= Gly, Val"
XX
PN WO9721823-A1.
XX
XX 19-JUN-1997.
XX
XX 07-NOV-1996; 96WO-JP03253.
XX
XX 12-DEC-1995; 95JP-0323285.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX Asada K, Kato I, Mitta M, Morishita M, Takakura H;
PI Tsunasa S, Yamamoto K;
XX WPI; 1997-332794/30.
DR P-PSDB; AAW24122.
XX
XX Protease(s) and genes encoding them obtained from *Thermococcus* and
PT *Pyrococcus* strains - have extremely high thermal stability and are
PT useful industrially and as research reagents
XX

Db 693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCA 752
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
 Db 753 CGTAGCTGGTATTGCGACCTCTTGCTCCAA-----GCACACCC 791
 QY 280 olus-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
 Db 792 GAGCTGGACTCCAGCAAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851
 QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrPheArgVa 310
 Db 852 AGATGAATAGCGGATATAGCTACGCTGCA-----GGTAGGGT 890
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSerThrSe 330
 Db 891 TAATGCATACAGGCTATAAC-----TACGATAACTATGCAAAAGCTAGTGTCTACTGG 944
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 Db 945 ATATGTTGCCAACAAAGGCGCAAACTCACCAAGTTCGTTATTAGCGGAGCTTCGTCGT 1004
 QY 345 ulysIleSerLeuValTyrSerAspAlaProAlaSerThrAlaSerValThrLeuVa 365
 Db 1005 AACTGCCACATTATCTGGGCAATGCCAAT----- 1035
 QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 Db 1036 -AGCGACCTTGATCTTACCTCTACCATCCCAATGGAACCCAG---GTTGACTACTCTTA 1091
 QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1092 CACCCCTACTAT-----GGATTGCAAAAGTTGGTTATTA 1127
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1128 CAACCCACTGATGGAACATGGCAATTAAGGTTGTAGCTACAGC-----GGAAG 1178
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1179 TGCAACTATCAAGTAGATGTGGTAAGT 1206

RESULT 10

AAT85695

ID AAT85695 standard; DNA; 1962 BP.

XX AC AAT85695;

XX DT 20-APR-1998 (first entry)

XX DE Pyrococcus furiosus PFUS protease coding sequence.

XX DE Protease; research reagent; thermal stability; pyrococcus furiosus; ss.

XX KW Pyrococcus furiosus DSM-3638.

XX OS WO9721823-A1.

XX PN 19-JUN-1997.

XX PP 07-NOV-1996; 96WO-JP03253.

XX PR 12-DEC-1995; 95JP-0323285.

XX XX (TAKI) TAKARA SHUZO CO LTD.

XX XX Asada K, Kato I, Mita M, Morishita M, Takakura H;

XX PI Tsunasa S, Yamamoto K;

XX XX WPI; 1997-332794/30.

XX DR P-PSDB; AAW24129.

XX XX

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

XX

SQ

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-985-689a-1 (1-434) x AAT85695 (1-1962)

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

Protease(s) and genes encoding them obtained from Thermococcus and
 Pyrococcus strains - have extremely high thermal stability and are
 useful industrially and as research reagents

Disclosure; Page 123-125; 159pp; Japanese.

This sequence represents the coding sequence for the protease from
 Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
 invention. The proteases of the invention have extremely high thermal
 stability. The proteases can be used as research reagents, and
 industrially in the food, drug and chemical industries.

Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 other;

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-985-689a-1 (1-434) x AAT85695 (1-1962)

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCA 752

260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280

753 CGTAGCTGGTATTGCGACCTCTTGCTCCAA-----GCACACCC 791

280 olus-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293

792 GAGCTGGACTCCAGCAAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 851

294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrPheArgVa 310

852 AGATGAATAGCGGATATAGCTACGCTGCA-----GGTAGGGT 890

310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSerThrSe 330

891 TAATGCATACAGGCTATAAC-----TACGATAACTATGCAAAAGCTAGTGTCTACTGG 944

330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345

945 ATATGTTGCCAACAAAGGCGCAAACTCACCAAGTTCGTTATTAGCGGAGCTTCGTCGT 1004

345 ulysIleSerLeuValTyrSerAspAlaProAlaSerThrAlaSerValThrLeuVa 365

1005 AACTGCCACATTATCTGGGCAATGCCAAT----- 1035

365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385

1036 -AGCGACCTTGATCTTACCTCTACCATCCCAATGGAACCCAG---GTTGACTACTCTTA 1091

385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405

1092 CACCCCTACTAT-----GGATTGCAAAAGTTGGTTATTA 1127

405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425

1128 CAACCCACTGATGGAACATGGCAATTAAGGTTGTAGCTACAGC-----GGAAG 1178

425 oGlnThrPheSerLeuAlaIleValAsn 434

1179 TGCAACTATCAAGTAGATGTGGTAAGT 1206

RESULT 10

AAT85695

ID AAT85695 standard; DNA; 1962 BP.

XX AC AAT85695;

XX DT 20-APR-1998 (first entry)

XX DE Pyrococcus furiosus PFUS protease coding sequence.

XX DE Protease; research reagent; thermal stability; pyrococcus furiosus; ss.

XX KW Pyrococcus furiosus DSM-3638.

XX OS WO9721823-A1.

XX PN 19-JUN-1997.

XX PP 07-NOV-1996; 96WO-JP03253.

XX PR 12-DEC-1995; 95JP-0323285.

XX XX (TAKI) TAKARA SHUZO CO LTD.

XX XX Asada K, Kato I, Mita M, Morishita M, Takakura H;

XX PI Tsunasa S, Yamamoto K;

XX XX WPI; 1997-332794/30.

XX DR P-PSDB; AAW24129.

XX XX

Tue Jul 29 14:12:39 2003

us-09-985-689a-1.p2n.rng

```

1038 TGTGCTCCAGAACTGGATAATGCTGCCAGACCAAGT-----GGAACTAGCAT 1088
240 eTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro11 260
1089 GGTGCAACCAATTAATGACTATTACACAGCAGCTCTGGACATCAATGGCAACTCTCA 1148
260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLleThrPr 280
1149 CGTAGCTGGATTGTCAGCCCTCTTGCTCAA-----GCACACCC 1187
280 oLys-----ProSerLeuLeuLysAlaAlaLeuLleAlaGlyAla----- 293
1188 GAGCTGAGCTCCAGACAAGTAACAGCCCTCATAGAACTGCTGATATCGTAAGCC 1247
294 -----AlaAspLleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
1248 AGATGAATAGCCGATATAGCTACGGTGCA-----GGTAGGGT 1286
310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSerThrSe 330
1287 TAATGCATACAAAGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTCTACTGG 1340
330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
1341 ATATGTTGCCAACAAAGCCAGCCAACTCACCAGTTCGTTATTAGCGGAGCTTCGTCGT 1400
345 uLysLleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
1401 AACTGCCACATATATCTGGACAATGCCAAT----- 1431
365 lAsnAspLeuAspLeuValLleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
1432 -AGCCACCTTGATCTTACCTCTACGATCCCAATGGAACACGAG---GTTGACTACTCTTA 1487
385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
1488 CACCGCCCTACTAT-----GGATTGGAAGAAGGTGCTGTTATTA 1523
405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
1524 CAACCCACTGATGAACATGGAACATTAAGGTTGTAAGCTACAGC-----GGAAG 1574
425 oGlnThrPheSerLeuAlaIleValAsn 434
1575 TCCAAACTATCAAGTAGATGTGTAAGT 1602

```

RESULT 11

ID AAX05929
AC AAX05929 standard; DNA; 1962 BP.

XX AAX05929;
XX AAX05929;

DT 06-MAY-1999 (first entry)

DE Hyperthermostable protease encoding DNA.

XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
XX Pyrococcus furiosus.

OS W09856926-A1.

PN 17-DEC-1998.

XX 04-JUN-1998; 98WO-JP02465.

XX 10-JUN-1997; 97JP-0151969.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Asada K, Kato I, Morishita M, Shimojo T, Takakura H;

XX WPI: 1999-080907/07.
DR P-PSDB; AAW94841.
XX Recombinant hyperthermostable protease from *Pyrococcus furiosus* -
PT and gene encoding it, for large scale production of the protease for
XX industrial use.
XX Disclosure; Page 59-60; 82pp; Japanese.
XX The invention relates to a hyperthermostable protease derived from a
CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
CC (optimum 6-8), and retains more than 90% of its activity after 8 hours
CC at 95 deg.C. The invention also provides gene sequences encoding a
CC polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal
CC peptide from subtilisin, and PRO is the above protease. Host cells
CC (especially *Bacillus* strains) transformed with vectors comprising the
CC genes are used for the recombinant production of the protease. The
CC hyperthermostable protease which can be prepared in quantity suitable
CC for industrial use, can be used as an additive for drugs, washing agents
CC and foodstuffs and for chemical synthesis.
XX Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 other;

Alignment Scores: 6.77e-23 Length: 1962
Pred. NO.: 416.50 Matches: 141
Score: 42.55% Conservative: 59
Percent Similarity: 30.00% Mismatches: 148
Best Local Similarity: 30.00% Indels: 122
Query Match: 20 Gaps: 20

US-09-985-689A-1 (1-434) x AAX05929 (1-1962)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlygl 24
Db 420 GTCGCGAGCTCAAGTTATGGCACTTACGTTGGAACTGGGATATGATGTTCTGGAAT 479
QY 24 nleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisgl 44
Db 480 CACAATAGGAATAATTGACACTGGAAATGAC-----GCTTCATCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAGGAAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 566
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGGAGTTATCCATACGATACCATGACATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGCTCCAGGAGCTAAGCTGCGGG 686
QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAGGTTCTAGTCCCGATGCTCTGGAAGCATATCTACTATAATTAAGGGAGTTGA 746
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGGCGGTGATACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTTCTGTTTC 806
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGCTACTGAGCTCTAAGTCAGGCTGTTAATGACGCTGGGATGC 866
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTGTGTGTGTGTCGCCG 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180

Tue Jul 29 14:12:39 2003

us-09-985-689a-1.p2n.rng

Thermoanaerobacter yonsei.

Key Location/Qualifiers
CDS 142..1779
/*tag= a
/product= "subtilisin-like serine protease"

OS

XX

FT

FT

FT

XX

PN

XX

PD

XX

PF

XX

PR

XX

PA

XX

PI

XX

DR

DR

XX

PT

XX

PS

XX

CC

CC

XX

SQ

Alignment Scores:

Pred. No.: 2,01e-19

Score: 372.00

Percent Similarity: 46.29%

Best Local Similarity: 33.25%

Query Match: 16.56%

DB: 23

US-09-985-689a-1 (1-434) x ABL54900 (1-2121)

QY

6

GlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlnLeu

25

Db

445

GGATACACAAA

26

ValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAla

45

Db

496

ATACCAATTATGACACAGGTATAGACGGAAATCACGTTGACCTCTCA

543

QY

46

PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn

63

Db

544

---GGTGGAAAATA

QY

64

---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn

79

Db

586

ACTACACCATACGACACAAATGCCATGGAATGACAGTATTGCTGCAGGTACA

645

QY

80

GlySerThrAsn

96

Db

646

GGTCTGGAAACAGCTTTTACAAAGCGTGTCTCTGATGCTTTGTTGGTAGGAATAAAA

705

QY

97

IleMetAspSerGlyGlyGly

110

Db

706

GTCTTAGATGCAATGGAAGCGGACATGAGCAGTACTGAGGAATTTGACTGGCT

765

QY

111

LeuGln

118

Db

766

GTTCAAATAAAGATGATATACGAATCAAAATTAATAAATTAAGCCCTGGCAGCTTCTACA

825

QY

119

TyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyr

138

Db

826

AGTTCTGATGA

864

160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI

180

891 CGGCAACAGCGCGCACACCTACACCGTCGCTCACCGCGCGCGAGCAAGGTAT

950

180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs

200

951 AACCGTCGGTGCA

200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa

220

978 CAACATCGCGCTCTCCAGCAGGACCGACCGCGGACGAGGCTCAAGCGGAAGT

1037

220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh

240

1038 CGTCGCCCGCGGTGACATATAGCCGCGCGCGCAGC

1088

240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI

260

1089 GGGCACCCCGATAAACGACTACTACACCAAGGCTCTGGAACCGACGATGGCCACCCGCA

1148

260 eValAlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleTh

279

1149 CGTTTCGGCGTGGCGCTCATCTCCAGGCCAC

1185

279 rProLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla

293

1186 -CCGAGCTGGACCCCGCACAAAGGTGAAGACCGCCTCATCGAGACCGCGACATAGTCGC

1244

294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyAr

309

1245 CCCCAGGAGATAGCGACATCGCTACGCTGGCG

1283

309 gValThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSerSerSe

326

1284 GGTGAACGCTACAAAGCCATCAAGTACGACGACTACGCCAAGCTCACCTTCACCGGCTC

1343

326 rLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeuLy

346

1344 CGTCGCCGACAAAGGAGCGCCACCTTCGACGTCAGCGCGCCACCTTCGTGAC

1403

346 sIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAs

366

1404 CGCCACCTCTACTGGAC

1433

366 nAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPheTh

386

1434 CGACATCGACCTCTACTCTACGACCCCGCACCGGGAACGAG---GTTGACTACTCTACAC

1490

386 rSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAl

406

1491 CGCCTACTAC

1526

406 aProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyProGl

426

1527 CCGACCGCGGAAACCTGGACGGTCAAGGTGCTAGCTACAG

1577

426 nThrPheSerLeuAlaIleValAsn

434

1578 GAATACACAGGTGCGTCGTCGTCAGC

1602

RESULT 13

ABL54900

ID ABL54900 standard; DNA; 2121 BP.

XX

AC ABL54900;

DT 31-MAY-2002 (first entry)

XX

XX T. yonsei subtilisin-like serine protease coding sequence.

XX Subtilisin-like serine protease; ss.

KW

Db	1172	CAGTCCAAAGGCAAGTACAAGGGCGTTCGACCCCGCGCGCATCTCTCAACGGCAAGGTC	1231
QY	98	MetaspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGln	117
Db	1232	CTCAGCGACTCCGGT-----TTCGGCGACGACTCCGGCATCTCTCGCGGCATGGAGTGG	1285
QY	118	AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTyrGlyAlaAlaValAsnGlyAl	137
Db	1286	GGCGCGCGGAGGCGCGGACGTCGTCACATGAGCTGGCGGCATGGACACACCGGAG	1345
QY	137	aTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLe	157
Db	1346	ACGACCCGCTGGAGGCGCGG-GTGCAACAAGCTGTCGCGCGAGAGGCGCTCTTCGCG	1404
QY	157	uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy	177
Db	1405	CATCGCGCGCGCAACAGGCGCGGAG-----TCGATCGTTCGCCCGGACGCGCGGA	1455
QY	177	sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs	197
Db	1459	CGCCGCGCTCACCGTTCGGCGCC-----GTGCA	1485
QY	197	pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro--ThrLysAspGlyArgIl	216
Db	1486	CGACAAGGACACAGCTCGCGACTTCCTCCACCGCGCCCGCTCGCGGACGCGGCAT	1545
QY	216	eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr	236
Db	1546	CAAGCGGACGTCAACGCTCCCGGCGTGACATCACGGCGCTCGCGGAGGCAACGA	1605
QY	236	oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe	256
Db	1606	CATCGCGCAGGAGTGGTGAGGACCGCGCGGCTACATGACCATCTCCGCGACGTGCAT	1665
QY	256	tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr	276
Db	1666	GGCGACCCCGACGTCGCGGCGCGCGCCCTCTGAACGACGAG-----	1711
QY	276	gClyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl	293
Db	1712	-----CACCCGACGTGACCTTCGCCGGAACCTGAAGGGCGCGCTCACCGGCTC	1758
QY	293	aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs	313
Db	1759	CACCAAG--GGCGCAAGTACACCCCGTTCAGACGAGGTTGCGGCGCGATCCAGGCCGA	1815
QY	313	pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer-----	328
Db	1816	CAAGGCGCTCAGCAGACCGGTATCGCGACCCGCTCGGTGAGCTTCGGCGCTCCAGCA	1875
QY	329	-----ThrSerGlnLysAlaThrThrThrSerPheThrAlaTh	340
Db	1876	GTGCGCGCACCGACGACGCGGTCAACGACGATGACCTACCGCAACCTCGGCGAC	1935
QY	340	rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp-----	353
Db	1936	CCAGGACGTACGCTGAAGCTGACGTGACCGCCACCGACCCCAAGGGCAAGCGGCCCC	1995
QY	354	-----AlaProAlaSerThrThrAlaSe	361
Db	1996	GGCGGGCTTCTTCGCTGGCGGCACACCGGTACCGCTCCGCGGCGCGGACGCGCTC	2055
QY	361	rValThrLeuValAsnAspLeuAspLeu-----	371
Db	2056	CGTTCGACATGACCGCGCACCCCGCTCGGCGGCACGGTGGACGGCGGTACTCGCGGTA	2115
QY	371	lIleThrAlaProAsnGlyThrGlnTyrVal	381
Db	2116	CGTGTGTCGACGGGCGCGGCGGACGAGTGC	2146

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: July 28, 2003, 02:45:54 ; Search time 5415.74 Seconds
(without alignments)
3278.366 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVOAYNPVGPQTFLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q/cygn2.1/USPRO.spool/US09985689/runat_25072003_143032_832/app_query.fasta_1.1166
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPC1=0 -LOOPEXP=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 -CGCN_1_1_5962@runat_25072003_143032_832 -NCPU=6 -ICPU=3
-NO_MMAR -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_fod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2247	100.0	1923	1	AB051423
2	2183	97.2	1920	1	AB046403
3	2143	95.4	1302	1	AB046406
4	2125.5	94.6	3003	6	AR069954
5	2125.5	94.6	3003	6	BD062155
6	1998.5	88.9	1299	1	AB046405
7	1994.5	88.8	1299	1	AB046402
8	1987.5	88.5	1299	1	AB046404
9	1986.5	88.4	2218	6	E03808
10	632.5	28.1	60006	1	AF268611
11	523.5	23.3	5890	3	DDU60086
12	511.5	22.8	5772	3	AF466309
13	497	22.1	6115	3	DDU20432
14	452.5	20.1	1977	6	AR201152
15	447	19.9	5109	3	AF263455
16	425.5	18.9	126928	2	AC096673
17	416.5	18.5	1236	6	AR201146
18	416.5	18.5	1962	6	AR201155
19	416.5	18.5	12452	1	AE010265
20	406.5	18.1	14376	1	AE013049
21	398	17.7	303450	1	SC0939130
22	381	17.0	299850	1	AP004601
23	366	16.3	3348	1	D83672
24	363.5	16.2	300425	1	AP005044
25	351	15.6	2539	6	AR202322
26	349	15.5	291000	1	SC0939105
27	347.5	15.5	292550	1	AP001513
28	346.5	15.4	302300	1	AP005034
29	346	15.4	2809	6	AR202321
30	346	15.4	3900	1	AB007809
31	338	15.0	1239	1	AF305633
32	338	15.0	1239	1	AY028704
33	338	15.0	135638	1	AF484556
34	337	15.0	12295	1	AE013026
35	332.5	14.8	1329	6	AX433519
36	331	14.7	233780	1	BSUR0010
37	328.5	14.6	300800	1	SC0939112
38	327.5	14.6	301150	1	AP004602
39	324.5	14.4	6854	1	BSU39230
40	320.5	14.3	2731	1	AF015225
41	310.5	13.8	2931	1	BACVPR
42	310.5	13.8	97015	1	BSGENR
43	310.5	13.8	212150	1	BSUP0020
44	309.5	13.8	1520	1	BS238598
45	309	13.8	4197	1	AF239672

ALIGNMENTS

RESULT 1

AB051423
 LOCUS AB051423 1923 bp DNA linear BCT 10-MAY-2002
 DEFINITION Bacillus sp. KP43 PROF gene for protease, complete cds.
 ACCESSION AB051423
 VERSION AB051423.2 GI:20521154
 KEYWORDS
 SOURCE Bacillus sp. KSM-KP43
 ORGANISM Bacillus sp. KSM-KP43
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 Itoh, S. and Saeki, K.
 new protease
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 1923)
 REFERENCE Saeki, K.
 AUTHORS Direct Submission
 TITLE Submitted (21-NOV-2000) Katsuhisa Saeki, KAO, CORPORATION;
 JOURNAL 2606, AKABANE, ICHIKAWA, HAGA, TOCHIGI 321-3486, Japan
 (E-mail: 387185@kantanet.kao.co.jp, Tel: 81285687471 (ex. 7471),
 Fax: 81285687403)
 COMMENT On May 9, 2002 this sequence version replaced gi:14164344.
 FEATURES
 Location/Qualifiers
 1..1923
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="genomic DNA"
 /strain="KP43"
 /db_xref="taxon:109322"
 1..1923
 /gene="PROF"
 1..1923
 /gene="PROF"
 /codon_start=1
 /transl_table=1
 /product="protease"
 /protein_id="BAB53674.2"
 /db_xref="GI:20521155"
 /translation="MRKKKKVFLSVLSAAALSTVALSNPSAGGARNFLDFKGIOTT
 QLEKFGKQGGTGAALFVSENVKLPKGLQKLETVPANNKLHIQFNGPILEETKQ
 TDEKFGKILQYIPDYAVIEVEGVKFSATSTIEHVESVEPLPIYRIDPQLETKGAS
 ELKVALDTRKQKKEVOLRGIEIOAQAISNDVLYITAKPEYKVMNDVARGIVKADV
 AQSGLYGCQIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTNNANDNGHGT
 HVAQSLGSLGNTKMGAPQANLVFQSIIMDSGGGLGGLPNSLQTLFQSAQSAARIHTN
 SWGAAYNGATDTSRNDVYVRKNDMTILFAAGNEPGNGTISAPGTAKNAITVGTPE
 NLRPSFYADYNHVAQFSSRPGTKDGRIPKPDVMAQPTFILSARSLAPDSSFWANH
 DSKVAYMGCTSMATPIVAGNVAOLREHFKVNRGIPKPSLLKAALTAGAADIGLYPN
 GNQGWRYLQKSLNVAYNVSSLSLTSOKATYSFTATAGKPLKISLVSDAPASTA
 SVTLNDLDELITAFNGTQYVGNDFSTPYNDWNGNENVEFINAPQSGTYTIEVQA
 YNPVGPQTFSLAIYN"

source

gene

CDS

BASE COUNT 576 a 418 c 475 g 454 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,44e-133 Length: 1923
 Score: 2247.00 Matches: 434
 Percent Similarity: 100.00 Conservative: 0
 Best Local Similarity: 100.00 Mismatches: 0
 Query Match: 100.00 Indels: 0
 DB: 1 Gaps: 0

US-09-985-689A-1 (1-434) x AB051423 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 |||||
 Db 619 AATGATGTTGCGCGTGAATTTGCAAGCGGATGTGGCTCAGACGACGTACGGGTGTAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 |||||
 Db 679 GGACAAAGGACACATCTGCGGTTCGCGATACAGGCGCTTGATACAGTCCGAATGACAGT 738
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 |||||
 Db 739 TCGATGCATGAAGCCCTCCCGCGGAAAAATTAATGATGATGATGATGATGATGATGATGAT 798

Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 |||||
 Db 799 AATCCCAATGATACAAATGGTTCATGGTACGATGTGGTGGCTCCGTATATTAGGAACGCG 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 |||||
 Db 859 TCCACTAATAAAGGAATGCGCCTCAGGCGAATCTAGTCTTCCAATCTATCATGGATAGC 918
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 |||||
 Db 919 GGTGGGGACTTGGAGGACTACCTTCGAATCTGCAAACTTATTCAGCAAGCATACAGT 978
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
 |||||
 Db 979 GCTGGTCCAGAAATTCATAAATCTCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 |||||
 Db 1039 GATTCACAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCTTTTCGCTGCC 1098
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 |||||
 Db 1099 GGAATGAAGGACCGAAGCGCGAACCACATCAGTGCACCGACAGCTAAATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 |||||
 Db 1159 ACAGTCGGAGCTACGGAACCTCCGCCCAAGCTTTGGGTCTTTATGCGGAGCAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 |||||
 Db 1219 CATGTGGCACAGTCTCTTCACGTGGACCGAAGGATGGACGGATCAACCGGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 |||||
 Db 1279 ATGGCACCGGGAACGTTTACTATCAGCAAGATCTTCTCTTGACCGGATTCCTCCTTC 1338
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 |||||
 Db 1339 TGGGCGAACCATTCACAGTAGTAATATGATACATGGTGGACGTCCATGGCTACACCGATC 1398
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 |||||
 Db 1399 GTTGTGGAACCGTGGCACAGCTTCGTGACCAITTTGTGAAAAACAGAGGCATCACACCA 1458
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 |||||
 Db 1459 AAGCCTTCTCTATTAAGCGGCACGTGATTCGCGGTGCAGCTGACATCGGCTTGGCTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 |||||
 Db 1519 CCGAACGGTAAACCAAGGATGGGACGAGTGACATGGATAAATCCCTGAACGTCGCTAT 1578
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 |||||
 Db 1579 GTGAAGCAGTCCAGTTCTCTATCCACCAGCAAAAGGACGACTACTCTTTACTGCTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 |||||
 Db 1639 GCCGGCAAGCCTTTGAAAAATCTCCCTGTTGGTCTGATGCCCTTGGCAGCACACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 |||||
 Db 1699 TCCGTAACGCTGCTCAATGATCTGGACCTTGCTCATTCACCGTCCCAATGGCACACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
 |||||
 Db 1759 GTAGGAAATGACTTTTACTTCGCCATACAATGATAACTGGGATGGCGCAATACCGTAGAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 |||||
 Db 1819 AATGATATTATTATGACCCACAAAGCGGACCTATACAAATTGAGGTACAGGCTTATAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434

Dn	1879	GTACCGGTTGGACCAGACACTTCTCGTTGGCAATTGTGAAT	1920
RESULT 2		AB046403	
LOCUS		Bacillus sp. 9860 PROA gene for protease, complete cds.	
DEFINITION		AB046403.2 GI:20521152	
ACCESSION		Bacillus sp. 9860	
VERSION		Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
KEYWORDS		Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and	
SOURCE		Horiokoshi,K.	
ORGANISM		Novel oxidatively stable subtilisin-like serine proteases from	
REFERENCE		alkaliphilic Bacillus spp.: enzymatic properties, sequences, and	
AUTHORS		evolutionary relationships	
TITLE		Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)	
JOURNAL		20568675	
MEDLINE		11118284	
PUBMED		2 (bases 1 to 1920)	
REFERENCE		Saeki,K.	
AUTHORS		Direct Submission	
TITLE		Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,	
JOURNAL		Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,	
COMMENT		Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,	
FEATURES		Tel.:81-285-68-7400, Fax:81-285-68-7403)	
source		On May 9, 2002 this sequence version replaced gi:12381938.	
gene		Location/Qualifiers	
CDS		1..1920	
		/organism="Bacillus sp. 9860"	
		/mol_type="genomic DNA"	
		/strain="9860"	
		/db_xref="taxon:133778"	
		1..1920	
		/gene="PROA"	
		1..1920	
		/gene="PROA"	
		/codon_start=1	
		/transl_table=1	
		/product="protease"	
		/protein_id="BAB21266.2"	
		/db_xref="GI:20521153"	
		/translation="MRKKKVFSLVSAAILSTVALNPNPAGARTFDLDFKGITTT	
		DVSFSGRQRTGAAFTVESENKLLGLKLLETVPANNKLHIVQFNPLETKOK	
		LETTGAKILBYIPDYAVIVEYEGDVQSKVRSEIHVESPEYLPKYKDLPOLFTKGAST	
		LVKALALDTKONNKREVLRLIEETAOXVASNDVHYITAKPEYKVMDVARGIVRADVA	
		OSSGYLGOGCIIVADPTGLDITGRNDSMHFAFGKITIALYLRGTNNANDNGHGH	
		VAGSVLGNATNKMPQANLVFQSDINDSSGGLFLSNLTQLFSQAFSAGARIHTNS	
		WGAANGAYTTIDSRNVDDYYVRKNDMTILFAAGNERPNPGTISAPGTAKNAITVGATEN	
		LFPSFGADINIHVAQSFRRGPDKGRIKPDVMAPGTIILSARSLAPDSPFWANHD	
		SKYAVMGTSMTAPIVAGNYAQRUREHFVKNRGITPKPSLLKAALIAAADVGLGYPNG	
		NOWGRVTLDKSLNAVYNESALLSTSQKATYTFATAGPLKLSLWSDPAPTAS	
		VTLVNDLDLVIAPNGTRYGVNDIFSAPFDNWDNRNVENVFINSPSGTYTIEVQAY	
		NVPVGPQNFSLATVN"	
BASE COUNT	576 a	411 c	473 g 460 t
ORIGIN			
Alignment Scores:			
Pred. No.:	3,81e-129	Length:	1920
Score:	2183.00	Matches:	418
Percent Similarity:	99.31%	Conservative:	13
Best Local Similarity:	96.31%	Mismatches:	3
Query Match:	97.15%	Indels:	0
DB:	1	Gaps:	0
US-09-985-689A-1	(1-434) x AB046403	(1-1920)	
Qy	1	AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrrGlyLeuTyr	20
Dn	616	AATGATGTGCCAGAGTAITGTCAAGCCGATGTGCACAGCAGCACTACGCTTCAGCTT	675

QY 381 valGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 |||||.....
 Db 1756 GTCGGGAATGACTTCTCAGCACCATTGACAACTAATCGGATGGCGCAATAACGTAGAA 1815
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 |||||.....
 Db 1816 AATGATTTATTAATTCGCCCAAGTGGACATATACCAATGAGTGAACATATAAT 1875
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 |||||.....
 Db 1876 GTCCCGGTGGACACAAAACCTTCTCGTTGGCAATTGTGAAC 1917

RESULT 3

AB046406
 LOCUS AB046406 1302 bp DNA linear BCT 23-JAN-2001
 DEFINITION Bacillus sp. NV1 PROE gene for protease, partial cds.
 ACCESSION AB046406
 VERSION AB046406.1 GI:12381944
 KEYWORDS
 SOURCE Bacillus sp. NV1
 ORGANISM Bacillus sp. NV1
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (sites)
 AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and Horikoshi,K.
 TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
 JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 MEDLINE 20568675
 PUBMED 11118284
 REFERENCE 2 (bases 1 to 1302)
 AUTHORS Saeki,K.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp, Tel:81-285-68-7400, Fax:81-285-68-7403)
 FEATURES
 source
 1. .1302
 /organism="Bacillus sp. NV1"
 /mol_type="genomic DNA"
 /strain="NV1"
 /db_xref="taxon:133781"
 1. .1302
 /gene="PROE"
 <1. .>1302
 /gene="PROE"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB21269.1"
 /db_xref="GI:12381945"
 /translation="NDVARGIVKADVAQSSYGLYGQGVVAVDVTGLDTCGRNDSMHEAFRCIKITVALGRNTNANDPNCCHTHVAGSYLNGCTSNKGMAPOANLVFOSVMSDNGGLGSPNSVTLFQAYSGARIIHTNSMGAPVNGYTTDSRNVDVVRKNDMAVLFAAGVMPGNTILSAPGTAKNALIVGATENLRPSGVSADINLHVAFSSRPTKDGRIKPDVMPGPTLSRSLAPDSFWANHSRIATMGSTMAPIVAGNVQAQRHFIKRGITPKSLLKALITAGATDIDGLGYPGNGQGRVTLDKSLNVAFFVNETSSLSTNQKATYFTAQSGKPKILSLVSDAPASTASVTLDLIVITAPNGTKYVGNDFTPAYDNNWDGRNNVNVFVNAQSGTYTVEVQAYNVPGQPAFLAIVN"

gene

CDS

377 a 283 c 322 g 320 t
 Alignment Scores:
 Pred. No.: 7,95e-127 Length: 1302
 Score: 2143.00 Matches: 406
 Percent Similarity: 97.93% Conservative: 19
 Best Local Similarity: 93.55% Mismatches: 9
 Query Match: 95.37% Indels: 0
 DB: 1 Gaps: 0

BASE COUNT
ORIGIN

QY 1 AsnAspValAlaAraGlyIleValLysAlaAspValAlaGlnSerTyrGlyLeuTyr 20
 |||||.....
 Db 1 AACCATGTTGCCAGAGGCAATTGTAAGCCGATGTTGCCAGAGCAGCTATGGTTTATAT 60
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 |||||.....
 Db 61 GGGCAAGGCAAGTGGTTGCAGTAGCGGATACCGGACTGGATACAGGGCGTAATGACAGC 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 |||||.....
 Db 121 TCGATGATGAAGCGTTCGAGGAAAGATTACCGCGATATATCCCTTGGAAAGCAAAAC 180
 QY 61 AsnAlaAsnAspThrAsnGlnHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 |||||.....
 Db 181 AACCCCAATGATCCAAACGGACACGGGACGATGTTGCCGATCTGTTTATGGAACCGGT 240
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 |||||.....
 Db 241 ACTTCGAATAAAGGATGGCTCCACAAGCTAAGTTAGTTTTCCTTCAATCTGTATGGACAGC 300
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 |||||.....
 Db 301 AATGGCGGACTTGGCGGACTGCCTTCCAATGTAAGTACATATTACCGCAGGCATATAGT 360
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
 |||||.....
 Db 361 CGCGGTGCAGAAATCCATACGAACCTCATGGGAGCGCTGTAACAGGAGCGGTACACTACT 420
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 |||||.....
 Db 421 GATTCAGAAAGAGTAGACGATTATGTCGTAATAATGATATGCGGTTCTTTTTCGACGC 480
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 |||||.....
 Db 481 GGTAACGAAGGCCCGAAATGGCGGCACAATCAGTGCCTCTGGTACCGCGAAGATGCTATC 540
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 |||||.....
 Db 541 ACAGTAGGGGCAACAGAAACCTGCGCCCAAGCTTTGGATCTTATGCTGCAACATCAAT 600
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 |||||.....
 Db 601 CATGTAGCACAGCTTTCTCCCGCGGACCTACAAAGGATGCACCTATCAAAACCGGACGTA 660
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 |||||.....
 Db 661 ATGGCACCAGGAACATTTTATTTATCGGCAAGATCTCTTTGGCTCCGGACTCCTCATTC 720
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 |||||.....
 Db 721 TGGGCAACCATGACAGCAAAATATGCTTATATGGTGAACATCCATGGCGACACCGATT 780
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 |||||.....
 Db 781 GTAGCTGTGAACGTTGCACAGTTACGTGAACATTTTCATCAAAACAGAGGAATCACTCT 840
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 |||||.....
 Db 841 AAACCATCTCTCTGAAAGCAGCTCTTATTGCGCGGACCACTGATATCGGCTTGGCTAT 900
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 |||||.....
 Db 901 CCGAGTGGAAACCAAGGATGGGGAAGAGTAACATTGGACAAGTCACTTAATGTAGCTTTC 960
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 |||||.....
 Db 961 GTAAATGAAACAAAGCTCTTTATCTACTTAAACCAAAAGGCTACGTATTCATTACTCACAA 1020
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 |||||.....
 Db 1021 TCAGGCAACACCTTGAAGATTTCATTTGTTGGTCTGATGCACCGGCAAGTACTTCCGCA 1080

US-09-985-689A-1 (1-434) x AB046406 (1-1302)

```

QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1081 TCGGTTACATTCGTTGAATGATCTGGATCTGATGATTCAGCTCCAAATGGACAAGTAT 1140
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1141 GTTGGAAACGACCTTTACTGCTCCCTATGATAAATAACCTGGGATGGACGTAACAATGTAGAG 1200
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrThrIleGluValGlnAlaTyrAsn 420
Db 1201 AACGCTGTTATCAATGCTCGCAAGCGGAACGTATACAGTTGAGGTACAGGCTTACAAT 1260
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1261 GTTCCACAAGGCGCGAGGCGTTTCTTTGGCTATTGTGAAC 1302

RESULT 4
AR069954
LOCUS AR069954 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS Unknwn.
SOURCE Unknwn.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3003)
AUTHORS Sloma,A. and Christianson,L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES
    Location/Qualifiers
        1..3003
            /organism="unknown"
BASE COUNT 982 a 504 c 645 g 872 t
ORIGIN

Alignment Scores:
Pred. No.: 2 86e-125 Length: 3003
Score: 2125.50 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.53% Mismatches: 8
Query Match: 94.59% Indels: 1
DB: 6 Gaps: 1

US-09-985-689A-1 (1-434) x AR069954 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCGCGTGGCATTGTGAAAGCAGACGTCGCACAAAATAACTTTGGCTATAT 1529
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GCACAAGACAGATTGTAGCAGTTGCTGATCTGGGCTTGATACAGGAAGAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGCATGAAGCATTCGCGGTAAAGATTACCGCAGTATATGCGCAGTATATGCGCAGCAAGCAAT 1649
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGAAAT --- 1706
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTACAATAAAGGATGGCCAGCCAGCAACCAATCTAGTCTTTCAATCTATTATGATAGT 1766
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGGAGGCTGGGAGGACTACCTGCTAATCTACAAACATTATTTCAGTCAAGCATATAGT 1826
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 1827 GCTGGAGCGAAGATTACCAATTCATGGGGGCTCCAGTAAACGGTCCCTATACGACA 1886

```

```

QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATTTCTTTTTCGGGCC 1946
QY 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAATGAGGGACAGGTAGCGGTACAATCAGTGCACGAGACAGCAAAATATGCCGATT 2006
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGGCAACCCAAACCTTACGTCGAAGCTTCGGATCTTATGCGGTAATATTAAC 2066
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATTTCTTCACGAGGCTCTACTAGAGATGGAGCTATTAAAGCGGACGTC 2126
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 2127 ATGGCACCAGGTACGTATATCTCTCTGTAGATCATCATAGCTCCAGATTCCTCATTC 2186
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 2187 TGGCAAAACCATGATAGTAAATATGCTACATGGGTGGTACTTCTATGGCTACTCCAATT 2246
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 2247 GTAGCAGGTAAATGTTGCCAATTAAGGGAGCATTTTGTGAAAAATAGAGGGTAACTCCT 2306
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCTTCCCTTTTAAAGCTCTTAAATGTCAGGTGCTGCGGATGTTGGACTTGGCTTT 2366
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATGCTAAACCAAGGATGGGGAAGTAGCTTAGATAAAATCCCTAAATGTCGCATTT 2426
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 2427 GTGAATGAACGAGCCCTTTTATCAACAAGTCAAAAGCAACATATTCGTTTACGGCTCAA 2486
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 2487 GCTGTAAACCCCTTAAATAATATCACTTGTGTCAGATGCACAGGTAGCAGCAGCGCA 2546
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 2547 TCAGTAATTTAGTAATGATTAGACTTAGTAATCACTGCACCAAAATGGAACCTAAATAC 2606
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 2607 GTCGGAATGACCTTTACAGCACCGCTATGATACAACATTCGGATGGCAGAAACACCGUGGAA 2666
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGTGTTTATCAATGCTCTCAAAAGCGGAACGTATACAGTCAAGAGTCAGGCTTACAAT 2726
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAGTAAGTCGCAAAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 5
BD062155
LOCUS BD062155 3003 bp DNA linear PAT 27-AUG-2002
DEFINITION Nucleic acids encoding a polypeptide having protease activity.
ACCESSION BD062155
VERSION BD062155.1 GI:22607760
KEYWORDS JP 2001514529-A/39.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3003)

```


source

```

1. .1299
/organism="Bacillus sp. SD521"
/mol_type="genomic DNA"
/strain="SD521"
/db_xref="taxon:133780"
1. .1299
/feature="PROD"
<1. .>1299
/feature="PROD"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21268.1"
/db_xref="GI:12381943"

```

gene

CDS

```

BASE COUNT 409 a 245 c 298 g 347 t
ORIGIN
Alignment Scores:
Pred. No.: 1.08e-117 Length: 1299
Score: 1998.50 Matches: 381
Percent Similarity: 94.24% Conservatives: 28
Best Local Similarity: 87.79% Mismatches: 24
Query Match: 88.94% Indels: 1
DB: 1 Gaps: 1

```

US-09-985-689A-1 (1-434) x AB046405 (1-1299)

```

QY 1 AsnAspValAlaAArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 1 AATGATGTAGCAAGAGGAATAGTAAAGCAGCAGCTTGCCACAAAACAATACGGACTATAT 60
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 61 GGACAAGTCAAGTAGTGTAGTACAGGATACGGGTTAGATACAGGTCGTAAACGATAGT 120
QY 41 SerMetHisGluAlaPheArgGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 60
DB 121 TCTATGATGATGAAGCATCCGTTGGGAAAATACAGCTCTTTACGCGTTTACGGAAGAACTAAC 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 181 AATGCAATGATCCGAATGGGATGGTACGCGATGTAGCAGGTCTCTGTACTTGGTAAT--- 237
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 238 GCTTTAAATAAAGGAATGGCTCCCAAGCTAACCTAGCTTCCTCCATCTATTATGGATAGC 297
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 298 AGCGGAGGATAGGTGGATACCATCGAATTTGAATACGTTATTAGTCAAGCTTGGAAAT 357
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
DB 358 CTGGGGCTAGATTCATACCTCTCTGGGGTCTCCAGTAAATGGAGCTACACTGCT 417
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 418 AACTCGAGACAAGTGGATAGTAGTTCGAAATATGATATACAGCTACTTTTTCAGCA 477
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 478 GGTAATGAAGGTCTTAATTCAGGAACAATAGTCTCCAGGCACAGCGAAAATGCCATT 537
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 181

```

```

DB 538 ACGTCCGGCCGCAACGGAAAACTATGCCCCGAGCTTTCGGTTCAATTAGCAGATAACCCAAAT 597
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 598 CATATCGCACAAATTTTCATCAAGAGGAGCTACGAGAGATGGACGAATTAACACGACGTA 657
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 658 ACAGCTCTCTGGGAACATTTATTTATCAGCAGCTTCTCCTTAGCCCCAGACACTCTCGTTT 717
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
DB 718 TGGCGGAATTATACAGTAAGTATGCGTACATGGCGGTACCTCTATGGCGACACTATA 777
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 778 GTTCGGGGGAATGTCGCGCAATACCGCAGCATTTTATAAAAAATAGAGGAATTACACCT 837
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 838 AAACCTTCCCTTATAAAGCTGACCTTATCCTGGGGCTACTGATGTTGGTTTAGGATAT 897
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 898 CCAAGTGTGACCAAGGCTGGGGCGTCTTACTCTAGATAAATCGTTAAATGACGGTAT 957
QY 321 ValAsnGlnSerSerSerSerSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 958 GTCAATGAGCAACTGCTATGACACAGGACAAAAAGCAACGATATTCGTTCCAGCACAA 1017
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
DB 1018 CGCGGTAACCTTTAAAAATCTCCTTAGTAGACAGATGACCCCTGGGAAGTACAACCTGCA 1077
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
DB 1078 TCTATACACTAGTTAATGATTTAGATCTAGTCTAGTCTCCGATGGACAAAATAT 1137
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
DB 1138 GTAGGAATGATTTAGTTATCTCTATGATAAATACTGGGATGTTGCGCAACAATGTTGAG 1197
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1198 AAGCTATTATAAACGCTCCGCAATCTGGACAGTATACAAITGAGGTTCAACGCGTAAAT 1257
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1258 GTACCATCTGGCCACAGCGTTTCTCACTAGCTATCGTACAT 1299

```

RESULT 7

```

AB046402
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

```

AB046402
Bacillus sp. D6 PROA gene for protease, partial cds.
AB046402.1 GI:12381936
Bacillus sp. D6
Bacillus sp. D6
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
Saeiki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.
Novel oxidatively stable subtilisin-like serine proteases from evolutionary Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
20568675
11118284
2 (bases 1 to 1299)
Saeiki, K.
Direct Submission
Submitted (20-JUL-2000) Katsuhisa Saeiki, Kao corporation,

Biological Science Laboratory; Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail: 387185@ekastanet.kao.co.jp).
Tel: 81-285-68-7400, Fax: 81-285-68-7403)

FEATURES

Location/Qualifiers
1..1299
/organism="Bacillus sp. D6"
/mol_type="genomic DNA"
/strain="D6"
/db_xref="taxon:127889"
1..1299
/gene="PROA"
<1..>1299
/gene="PROA"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21265.1"
/db_xref="GI:12381937"
/translation="NDVARGIVKADVAQNNGYGGQVAVADTGLDGRNDSMHE
AFRGKITALYALGRITNNANDPNHGHGTHVAGSVLGNALNKGMAPOANLVFQSDSSGG
LGGPISNLNTLPSOAWNAGARTHTNSWGPVNGAYTANSROVDEVYRNNDMTVLFPAAG
NEGPNSTISARCTAKNAITVCATENYRPSGSIADPNHIAQFSRSGATROGRKPD
VTAPGTFILSRADLPSFWANINSKTYMGTSMTPTVAGNVQAOLREHFIKRG
ITPESIKALIALAGATDVGYPGSDQGWRTLDKLSNVAIVNEATLFTGQKATY
SFQQAQKPLKISLWTDAGSTTASYTLVNDLIDLITAPNGQKYVGNDFSEYEDNNW
DGRNNEVFNTPQSGTYTIEVQAYNVPSGPQRFSLAIVH"
BASE COUNT 413 a 243 c 298 g 345 t
ORIGIN

Alignment Scores:
Pred. No.: 1.92e-117 Length: 1299
Score: 1994.50 Matches: 380
Percent Similarity: 94.01% Conservative: 28
Best Local Similarity: 87.56% Mismatches: 25
Query Match: 88.76% Indels: 1
DB: 1 Gaps: 1

US-09-985-689a-1 (1-434) x AB046402 (1-1299)

QY 1 AsnAspValAlaAtrGlyLeuValLysAlaAspValAlaGlnSerTyrGlyLeuTyr 20
DB 1 AATGATGTAGCAAGAGGAAATAGTAAGACAGACGTTGCACAAACAAATACGGACTATAT 60
QY 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 61 GGACAAGGTCAAGTGTGACGTAGCGGATACGGGTTAGATACAGGTCGTAAACGATAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysLeuThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 121 TCTATGCAATGAGCAATCCGTGGGAAATACAGCTCTTTACCGCTTAGGAGAACTAAC 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 181 AATGCAATGATCCGAATGGGCATGTACGATGTAGTGGTCTGTGCTTGGTAAT-- 237
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 238 GCTTTAAATAAGGAATGCTCCGCAAGTAACTTAGTCTTCCAAATCTATTATGGATAGC 297
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 298 ACCGAGGATAGGAGGATACATCGAATTTAAATACGTTATTAGTCAAGCTTGGAT 357
QY 121 AlaGlyAlaAtrGlyHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
DB 358 GCTGGCGGTAGAAATTCATACTAATCTTGGGGGCCCGACAGTAAATGGAGGCTACACGCT 417
QY 141 AspSerArgAsnValAlaAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 418 AACTCGACACAAAGTGATGAGTATGTTTCGAAACAACTATGATGACGGTATTTTTCAGCT 477
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180

478 GGAATGAAGGACCTAACTCTCGAACAAATAGCGCTCCAGGGACAGCGAAAAATGCCATT 537
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 538 ACGTGGCGCAACGGAACACTACCGAACGTTTGGTTCATTCAGATAACCTAAT 597
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 598 CATATCGCACAAATTTTCATCGAGAGGAGCTACGAGAGATGGAGCAATTAACACAGCGTA 657
QY 221 MetAlaProGlyThrPheIleLeuSerAlaAtrGlySerSerLeuAlaProAspSerPhe 240
DB 658 ACAGCTCTCGAACATTTATATATACGACGCTCTCTTTAGCACGACTCTTCGTTT 717
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
DB 718 TGGGCAATTTATACAGTAAGTATGCTATATGGCGGTACCTCTATGGGACACCTATA 777
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 778 GTTGGGGGAATGTCGCGCAATTCGCGACATTTTATAAAATAGAGCAATACACCT 837
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 838 AAACCTCTCTTAATAAAGCTGCACATTCGCTGGGCTACTGATGTTGGTTTAGGATAT 897
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 898 CCAAGTGTGACCAAGCTGGGCGCTGTACTTTAGATAAATCGTTAAATGTAGCGTAT 957
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 958 GTCAATGAGCAACTGCATTAACACAGGACAAACGACGATTCGTTCCAAACGCAA 1017
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
DB 1018 GCGGGTAACCACTTAAATCTGTTAGTAGACACATGCACCTGGAAGTACAACGCA 1077
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
DB 1078 TCTTATACACTAGTTAATGATTTAGATCTAGTTATTACTGCTCGCAATGGACAAAATAT 1137
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
DB 1138 GTAGGTAATGATTTAGTTATCCTTATGATTAATTTGGATGGTGGCGCAACATGTTGAG 1197
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1198 AACGTATTTATAAACGCTCGCAATCTGGAACGATATACAAATTGAGTTCAAGCGTATAAC 1257
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1258 GTTCCATCTGGACACAGCGTTTCTACTAGTATCGTACAT 1299

RESULT 8
AB046404 1299 bp DNA linear BCT 23-JAN-2001
LOCUS Bacillus sp. Y PROC gene for protease, partial cds.
DEFINITION AB046404
ACCESSION AB046404.1 GI:12381940
VERSION
KEYWORDS
SOURCE
ORGANISM
Bacillus sp. Y
Bacillus sp. Y
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 (sites)
AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and Horikoshi,K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE 20568675
PUBMED 11118284

REFERENCE AUTHORS TITLE JOURNAL

2 (bases 1 to 1299)
Saeki, K.
Direct Submission
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
Location/Qualifiers
1. 1299

FEATURES source

/organism="Bacillus sp. y"
/mol_type="genomic DNA"
/strain="y"

gene
/db_xref="taxon:133779"
1. 1299

CDS
/gene="PROC"
<1..>1299
/gene="PROC"

/codon_start=1
/evidence=experimental
/transl_table=1
/product="protease"
/protein_id="BAB21267.1"
/db_xref="GI:12381941"

/translation="NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTRNDSSMHE
AFRGKITALYALGRNNSADPNHGHTVAGSVLGNALNKGMAPOANLVQSIMDSGG
LGLPSNLNLFQSNAGARLHTNSWAGAYTANSROVDYVRNNDMTVLFAG
NEGPNSTGISAPQNAITVGTATENYRPSFGSIADPNHIAQFSKRGATRDGRIKPD
VTAPGTFLSARSLLAPDSFWANYSKYVMGTSMTAPIVAGNVAOLREHFIKRRG
ITPKPSLKAALITAGATDVGLGYPNGDQGWRTVNLKSLNVAIVNCAATATGOKATY
SFOAGKPLKLSLVYTDAPGSTTASYTLVNLDLVITAPNGOKYVGNDFSVPYDNNW
DGRNVNVCFINAPSGYILIEVATNVPFGPQFSLAIVH"

BASE COUNT 404 a 245 c 295 g 355 t

ORIGIN

Alignment Scores:

Pred. No.: 5,33e-117 Length: 1299
Score: 1987.50 Matches: 379
Percent Similarity: 94.01% Conservative: 29
Best Local Similarity: 87.33% Mismatches: 25
Query Match: 88.45% Indels: 1
DB: 1 Gaps: 1

US-09-985-689A-1 (1-434) x AB046404 (1-1299)

Qy 1 AsnAspValAlaAArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGAGCAAGAGGATAGTAAAGCTGATGTGCACAAACAAATACGGATATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAGTCAAGTAGTTCAGTAGCGGCACACAGGCTTAGATACAGGCTAACGATAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaGlyArgThrAsn 60
Db 121 TCTATGATGATGAAGCATTCGCGGGAATATCACAGCTCTTACGCGCTAGGAAGACTAAT 180
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuLeuGlyAsnGly 80
Db 181 AATCGAGTAGTCGAATGGCATGGCACATGATGACAGGTCTCTGTACTTGGTAAT--- 237
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 238 GCCTTAAATAAAGAAATGGCTCCGCAAGCTAACCTAGCTTCCCAATCTATTATGGATAGC 297
Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 298 AGCGGAGGATAGTAGTGGCTTACCATCAACCTAAATACGTTATTAGTCAAGCTTGGAA 357
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 358 GCTGGAGCAAGATTCATACCTCTTGGGAGCGCCAGCAAGTAAATGGAGCGTACACTGCT 417
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160

Db 418 AACTCGAGACAAGTGGATGAATATGTTCCAAATGAATGATGACGGTACTTTTTCAGCT 477
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 478 GGTAAATGAAGTCCCTAAATTCAGGAACAATAGTCTCCAGGTACAGCGAAATGCTATT 537
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 538 ACGGTCGCGCAACGAAACTATCCCAAGCTTCGGITCGATAGCAGATAACCCAAAT 597
Qy 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 598 CATATGCACAATTTTCATCGAGAGAGCTAGAGGATGGAGCAATTAAGACCTGACGTA 657
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 658 ACAGCTCTCGGAACATTTATTTATCAGCACGTTCTCTCTAGCTCCAGACTCTTCGTT 717
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 718 TGGCGGAATTAACAGTAAATACGCTATATGCGCGGTACCTCCATGGCGACACCTATT 777
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 778 GTTCGAGGAAATGTCGCAATACGTAGCATTATATAAAAAATAGAGGTATTACTCCT 837
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 838 AAGCCTCTTTTAAATAAAGCTGCACCTTATCGCTGGTCTACTGATCTTGGTTAGGATAT 897
Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 898 CCTAATGCTGACCAAGGCTGGGCGGTGTACTCTAAATAAATCGTTAAATGTAGCGTAT 957
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 958 GTCAATGAAGCAACTGCATTAGCCACAGACAGCAAAAGCAACCTATTCTGTTCCAGCACAA 1017
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1018 CGCGGTAAACCTTTTAAATACTCGTTAGTAGGACAGATGCTCTCGGAAGTACAACTGCA 1077
Qy 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1078 TCTATACACTAGTATGATTTAGATCTAGTATTACTGCTCCGAATGGACAAATAT 1137
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1138 GTAGGAATGATTTAGTTATCTTATCTTATGATAATAAATCGGATGGTCCGCAACATGTTGAG 1197
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1198 AAGCTATTTATAACGCTCCGCAATCTGGAACGTATATAATAGTGTTCAGGTTCAAGCGTATAAT 1257
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1258 GTACCATCTGCGCCACAGCGCTTCTCACTAGCTATCGTACAT 1299

RESULT 9

E03808
LOCUS 2218 bp DNA linear PAT 29-SEP-1997
DEFINITION DNA encoding alkaline protease Ya.
E03808
ACCESSION E03808.1 GI:2172022
VERSION JP 1992197182-A/1.
KEYWORDS Bacillus sp.
SOURCE Bacillus sp.
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 2218)
AUTHORS Tobe, S., Odera, M. and Asai, Y.
TITLE DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE PROTEASE YA USING THE DNA

JOURNAL Patent: JP 1992197182-A 1 16-JUL-1992;

LION CORP

COMMENT OS Bacillus sp.
 PN JP 1992197182-A/1
 PD 16-JUL-1992
 PF 28-NOV-1990 JP 1990327110
 PI TOBE SEICHI, ODERA MOTOYASU, ASAI YOSHITO
 PC C12N15/57, C11D3/386, C12N9/54, (C12N15/57, C12R1:07), (C12N9/54, C12R1:07);
 PC C12R1:07;
 CC strandedness: Double;
 CC topology: linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain-Y strain;
 CC *source: clone-PUB8A;
 FH Key
 FH Location/Qualifiers
 CDS 218..2125
 /product= 'precursor of alkaline protease Ya',
 sig_peptide 218..2125
 mat_peptide 824..2122
 /product= 'precursor of alkaline protease Ya',
 Location/Qualifiers
 1..2218
 /organism= 'Bacillus sp.'
 /mol_type= 'genomic DNA'
 /db_xref= 'taxon:1409'
 707 a 376 c 481 g 654 t

FEATURES

Source

BASE COUNT 707 a 376 c 481 g 654 t

ORIGIN

Alignment Scores:

Pred. No.: 1 2e-116 Length: 2218
 Score: 1986.50 Matches: 379
 Percent Similarity: 94.01% Conservatives: 29
 Best Local Similarity: 87.33% Mismatches: 25
 Query Match: 88.41% Indels: 1
 DB: 6 Gaps: 1

US-09-985-689A-1 (1-434) x E03808 (1-2218)

QY 1 AsnAspValAlaAArgGlyLeuValLysAlaAspValAlaGlnSerSerTyGlyLeuTyr 20
 DB 824 AATGATGTAGCAGAGGGATAGTAAAGCTGATGTCACAAACAAATACGGATTATAT 883
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 884 GCACAGGTCACACTAGTTCGAGTACGCGACACAGCTTAGATACAGGTGCTAACGATAGT 943
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 944 TCTATGCATGAAGCATTCGCGGGAAATCACAGCTCTTTACGCGTTAGGAAGAACTAAT 1003
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 1004 AATGCGAGTATCCGAATGGCATGGCACACATGATGACGAGTCTGTTACTGGTAAAT --- 1060
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 1061 GCTTTAAATAAGAAATGGCTCCCAAGCTAACCTAGCTTCCCAATCTATTATGGATAGC 1120
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 DB 1121 AGCGGAGGATAGTGGCTTACCATCGAACTAAATAGTATTATAGTCAAGCTTGGAAAT 1180
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 DB 1181 GCTGGAGCAGAAATCTACTACTACTCTTGGGAGCCCAAGTAAATGGACGCTACACTGCT 1240
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 1241 AACTCGAGACAAAGTGGATGAGTATCTTCGAAATAATGATATGACGCTACTTTTTCAGCT 1300
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180

DB 1301 GGTATGAAGTCCTAATTCAGGAACAATAGTCTCCAGGTACAGCAAAAATGCTATT 1360
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsn 200
 DB 1361 ACGGTCGCGCAACGAGAAATATCGCCCAAGCTTCGGTTTCGATAGCAGATAACCCAAAT 1420
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 DB 1421 CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGGATGGACCAATTAACCTGACGTA 1480
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 DB 1481 ACAGCTCTCTGGAACTTTATTTATCAGCACGCTTCTTCTTAGCTCCAGACTCTTCGTTT 1540
 QY 241 TrpAlaAsnHisAspSerLysTyAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 DB 1541 TGGCGCAATTAACAGTAACGCGTATATGGCGGTACTCCATATAAATAGAGGTATTACTCT 1600
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 DB 1601 GTTGAGGGAATGTCGCGCAATACGTGAGCATTTTATAAATAATAGAGGTATTACTCTCT 1660
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 DB 1661 AAGCCTTCTTTAATAAAGCTGCACCTTATCGCTGCTACTGATGTTGGTTTAGATAT 1720
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 DB 1721 CCTAGTGGTGACCAAGCTGGCGCTGTTACTCTAGATAAATCGTTAAATGACGTTAT 1780
 QY 321 ValAsnGluSerSerSerLeuSerThrSerglnLysAlaThrTyrSerPheThrAlaThr 340
 DB 1781 GTCAATGAAGCAACTGCATTAGCCACAGGACAAAAGCAAGCTATTTCGTCACAGCACAA 1840
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 DB 1841 GCGGGTAAACCTTTAAATAATCTCGTTAGTAGGACAGATGCTCCTCGGAAGTACAACTGCA 1900
 QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 DB 1901 TCTATACACTAGTTAATGATTTAGATCTAGTTATTACTGCTCCGAATGGACAAAATAT 1960
 QY 381 ValGlyAsnAspPheThrSerProTyAsnAspAsnTrpAspGlyArgAsnValGlu 400
 DB 1961 GTAGGAATGATTTTAGTTATCTTATGATATAAATACTGGATGGATGCTCGCAACATGTTGAG 2020
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 DB 2021 AACGTATTTATAAAGCTCCGCAATCTGGAAGCTATATAAATGAGGTTCAAGCGTAAAT 2080
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 2081 GTACCATCTGCGCCACACGCTTCTCTACTAGTATCGTACAT 2122

RESULT 10

AF268611/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

6006 bp DNA linear BCT 30-MAY-2001
 Uncultured marine group II euryarchaeote EBAC37Fl1, BAC clone
 37Fl1, partial sequence.
 AF268611
 AF268611.1 GI:9664575
 uncultured marine group II euryarchaeote 37Fl1
 uncultured marine group II euryarchaeote 37Fl1
 Archaea; Euryarchaeota; Marine Group II; environmental samples.
 1 (bases 1 to 6006)
 Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
 Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
 Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
 Construction and analysis of bacterial artificial chromosome
 libraries from a marine microbial assemblage
 Environ. Microbiol. 2 (5), 516-529 (2000)

```

MEDLINE      21128653
REFERENCE    11233160
AUTHORS      2 (bases 1 to 60006)
              Beja,O., Suzuki,M.T., Koonin,E.V., Aravind,L., Hadd,A.,
              Nguyen,L.P., Villacorta,R., Amjadi,M., Garrigues,C.,
              Jovanovich,S.B., Feldman,R.A. and Delong,E.F.
TITLE        Direct Submission
JOURNAL      Submitted (17-MAY-2000) R & D, Monterey Bay Aquarium Research
              Institute, P.O. Box 628, Moss Landing, CA 95039-0628, USA
FEATURES     Location/Qualifiers
              1..60006
              /organism="uncultured marine group II euryarchaeote 37F11"
              /mol_type="genomic DNA"
              /db_xref="taxon:133822"
              /clone="BAC 37F11"
              /note="environmental sample"
              1484..3178
              /note="37F11#1"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="S-layer protein"
              /protein_id="AAF97179.1"
              /db_xref="GI:9664576"
              /translation="MEPCQTELIDIFVQSPSAIAQVNIKIRVSDSGQGLEFEI
              PVVSPNPNELNASQSFVSEGGYPLAWNTGNTMSHVTITLENLPVGNISGMO
              SASUSPHEARGIPFEIIPPTNMDKQPLEIELSIVHPOLGMXVESITVIHNSISFATP
              VLSGASSQAFSLHANSINPTVTPDATNQDEYRLILEEGINFETISSNTDAIQP
              VNIGSLPQVSAQCFTSDVFOELGRAPLTETVYSCSVNGDLDEOTRLFTFASNTWGE
              TIPDLQFTIORNESRVNISVDQWDEPQGLLIRVVCYDNGLEVIESIAQVSRQ
              SNWVNGTASFSTGDLIDIAIRTVNEVLGDVNCILTVTSSTSSYRIERIVDEGSGQA
              PIYKIQNPSISDRSLSATIDCNSPFVDDDDPDSNTATAIYEESSQSLVTTNNLLWGT
              AETIFVAGVLMORREAGVLQENIRSKPKPTNNKDTQETKEVIEDDSLEIEE
              EMILEEPVSLVEITVSETDLSPLGRIDTRQELDPDVEIVDITTSIEERNKFFD"
              3223..3891
              /note="37F11#2"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="unknown"
              /protein_id="AAF97180.1"
              /db_xref="GI:9664577"
              /translation="MNQNSNSTAFYTLDCGVSENLDETQALIEIRGNHQLMEKLV
              NREAFSTRKKQGLDEIDRKNKTWDDATLWFTCVRMNLNDSAPIMLGKNKRRFGR
              FPHHTSGLOYILENLSIFSQSEGLYSDLIEHLEYLTERCSEMTGKHELEGKAGL
              MILGYITFEVKAURLSLGSGMWVSHEEPLDGVREAVRHLNALLMAAERRRGLIHH
              ROHE"
              complement(3892..4797)
              /note="37F11#3"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="SAICAR synthase"
              /protein_id="AAF97181.1"
              /db_xref="GI:9664578"
              /translation="MGEMLYQGVKQVWSTDDPDILFEFRNTNQISVEDQILPSLPRK
              GSNLRTTAHWKLVQEGVCGTHVELNAPDCLVKVKEVREKPAIPDAEWVFP
              BELVPHVYLSGNWRFPQGLFAESLGSODAEYGVKLESEVETTKFEAFDRNID
              REALQISNITDOEFSIVOAALKVDEIIEERAANKGLIHVDGKKEFALGPRGRVILV
              DFPTGLDNRWDADAAYSNGECIELSKFEVFRTHYINSQHAQLKLARDSGAIIDPPIA
              LPESVLEETASLYSSMYERLTSGQF"
              complement(4846..5508)
              /note="37F11#4"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="translation initiation factor 6"
              /protein_id="AAF97182.1"
              /db_xref="GI:9664579"
              /translation="MPIDFLDTHGSDQIGIHILASGVNVLFPHPRELTPPTIEKLEQVVG
              LELAFISGCSNLGALAGNKGMAVADIATESIDALTATYGDVVVMSEGVTAGNL
              LLVNQGGCVASPSIPKDGDLILSEVLGVDIISTIGQDQVGVSLGVNMGVLLHPDV
              APEVLLTEILGLVPPMVGTVSFGSPYVGAGICASNNGALAGSETTGPELNRIEDALG
              LI"
              5597..5745
              /product="5S ribosomal RNA"
              5816..8611
              /product="23S ribosomal RNA"
              complement(8888..10192)
              /note="37F11#7: contains Zn-ribbon and TPR-like repeats"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="unknown"
              /protein_id="AAF97183.1"
              /db_xref="GI:9664580"
              /translation="MNECAVCGILYLAGPSCGACGSQLRAOQSTSDSAAALPTEVPG
              LDDAAQAWYDLFEGIEAPVEQEEAASSLPFGFGESOTNISRPLFCIGSHPGICPEL
              SNDSGOEEVEVQHPVPIESTNLPEVVALETQHDSDPVVASVSEIPEVLPAPVAPD
              LPLIEPPVVPVPGPRLRLTAIPEVEPTVDTDLNPENWSIESFGSDQILFTAEVDIE
              VFSLEDLTVVHVEHDGENYPOMEKDSYSEKSGQPPDLHPAQLMVDVGTDSASRAL
              LESGFTSIGQDWRSAARSPQIVARFPQDSGAMNNYGIALQLQVASHVQHSNDPIEVS
              NSATOFEAAILSLREAVNTSSQPEIYNLSQALLSGREEKALALLDMAGNESQRES
              NFTNLRAAILAQUGRYDEAKRLVLPLOADKLASGNLLKLPAI"
              complement(10248..11969)
              /note="37F11#8"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="permease"
              /protein_id="AAF97184.1"
              /db_xref="GI:9664581"
              /translation="MGCNILLFKRHPAQLIALISLLFVGTMQFTIRSSHAAGGGDIDL
              TVTIRDFENQAVYVGLIGLLFYVALLITEVVRHTLAATLGLLAIATIALNHYSEAL
              SLKAVTMDWETIGLLGMVMVGIISHTGVFEWFAVQAYKSGGETATLWILCSV
              TAVLSAFLDNVTTMLLTPTVIKIAQVLDLKPILILISEVLFSNGAATMGDPNI
              MGSAMDPAISEGNKDNFIANLASDGVTFNDFITEMAPGILMTVPSEMLIRWLYKDE
              FSGRRDRDTELEQXAIKDKGLKVSAGAILTLVLCFELHPVHLAVSWALSGALV
              MLLATNRHLEPELHEVHTTLLFAGLVLVHLSLOYMGVTDYIGEVYKAIKFPDD
              YRLAAAILILWVSATASAFIDNIPYATMIPVLSLAFELNLPNLIALAFGACL
              GGNLTIGASNVVTAGMSEAGYPSISNEFFRAGFPVMLLSTFVSYFMILVYVVG
              EGALTWKIVLVGILGIIIAQYSRGRAGKSPAEALVDDDFEIVSTIKSVVITGS
              KGESEE"
              complement(12019..12282)
              /note="37F11#9: contains KH domain"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="RNA-binding protein"
              /protein_id="AAF97186.1"
              /db_xref="GI:9664583"
              /translation="MSVPNHIIHQAKDTSLPITMRIGKSGLTDSVVVELEGQLSSRL
              VKVKNRGLYPRDELKTVNMAEKTNSVVFARGNVAVFWRN"
              complement(12279..12614)
              /note="37F11#10: contains Zn-finger; conserved protein"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="unknown"
              /protein_id="AAF97185.1"
              /db_xref="GI:9664582"
              /translation="MGRRLNRDRKKGHAIWSTOLLKKNPSOFTPQORSREALH
              LVKVSRRIRILRPSREIMRCNTLLRYGENATIFRNGHKIQTCHSCTSVRRIPY
              RQHSQKVA"
              complement(12624..13274)
              /note="37F11#11"
              /codon_start=1
              /evidence=not_experimental
              /transl_table=11
              /product="adenylate kinase"
              /protein_id="AAF97187.1"
              /db_xref="GI:9664584"
              /translation="MDEGEROGLNVLFGPPGACKGTQAOHRMDATGLPQVSTGDML
              RAAKAGTSVGLKXSKMDKGAIVSDSIIIDIRLHDDDDAKNGVMFGFRTVPOA
              KALSEIVEVSAVLDAIDPDERIVERICGTCGTGVYHDTFNPIQNGICGGSFTE
              KRRADDVEETVLARLEAYHTQTSPLADEYRAAGIFHHVDGDRDIDVIDITDLLALVLC"

```

[illegible]

[illegible]


```

Db      1395  GGTACACATGTTTGGTCTTGCAGCAGGTGCACAGAGGATTCCTTCATTGGCTATTTC 1454
Qy      82   ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle---MetAspSer 100
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      1455  TCATTAGTGTCTTGGCACTGATGCAAGATTCGATTCCTTTGATTTAGCATCCGACCCA 1514
Qy      101  GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaThrSer 120
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      1515  AGTAATAATGAACCAAGTTCACCAAGATTTATATACCAATATATATACGCCATATATAAT 1574
Qy      121  AlaglyAlaArgIleHisThrAsnSerTrpGlyAla-----AlaValAsnGly 136
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      1575  GCAGTGCAGAGAGTACATGGTGTCTTGGGGTCTCTCTTCAATACAAAGGTATTTGGGT 1634
Qy      137  AlaThrThrThrAspSerArgAsnValAspAspTyrValArgLysAsn---AspMetThr 155
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      1635  AGTTATTCGAGATGATGCTGGTGGTATGATGATTCCTTTACATCCTCCAGATTTCAT 1694
Qy      156  IleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      1695  ATCTTTAGAGCTGCTGTGTAATAATGAACAATATATCATCATTCGTATCA---CAAGCAACA 1751
Qy      176  AlalysAsnAlaIleThrValGlyValAlaThrGlnAsnLeuArgProSerPhe----- 192
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      1752  GCCAAAAATGTAATACAGTTGGTCCGCAACAACACTACTCAGAAAGTTATACCACTGAT 1811
Qy      192  ----- 192
Db      1812  GCATTGGAATPATTCAATTTTGAACAGTAGCAAGAGATACATTAATTCATTATGCCAA 1871
Qy      192  ----- 192
Db      1872  TCATTTGATGATAAATATGTACATATACAACTGCTCAATGTTGTACAGAAATATCAACT 1931
Qy      193  -----GlySerTyrAlaAsp----- 197
Db      1932  GTTAAAGTTTATCAGGCTGTTGTACATCTTATATTAATAATTCATACGTTTCAATATT 1991
Qy      198  -----AsnIleAsnHisValAlaGlnPheSerSerArgGlyPro 210
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      1992  TCTCACAACCAAGATATATAATAGATAATATTTGTTTCATTCCTCTTCAAAAGGTCCA 2051
Qy      211  ThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAla 230
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      2052  ACACATGATGGTAGATTGAACCTGATATAGTTGCCCTGCTCAATATATATACATCGCA 2111
Qy      231  ArgSerSerLeuAla-----ProAspSerSer 239
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      2112  AGATCAAAATGGTGCAAAATACAAACCGATCAGTGTGTGATGGTCTTTTACCAATACAAAT 2171
Qy      240  PheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro 259
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      2172  GCACATATATACAGAA-----TCTGGTACATCAATGGCAACACCA 2210
Qy      260  IleValAlaGlyAsnValAlaGlnLeuArgGluHis----- 271
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      2211  TTAGCAACAGCAGCAACAACAATCTTAGACAATTTAGTTGATGTTATATATCAACT 2270
Qy      272  -----PheValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaLeu 289
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      2271  GGTCAATGTGAGAAATCAAAATTAACAACCAACTGGATCATTTATTAAGCATTAATG 2330
Qy      290  IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrPro----- 301
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      2331  ATTAATATGCTCAGTTTAAATGGTACATTCCTCCATTCATCATCAACAATAACAATCCA 2390
Qy      302  -----AsnGlyAsnGlnGlyTrpGlyArgVal 310
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      2391  TCAATGCAGATTATTGATACATTTGCAGGTGCAAAATTTGTTCAAGGTGGGTTTCATT 2450
Qy      311  ThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSer----- 325
      ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
Db      2451  AGAATGAGTGAATGGTTA-----TATGTTGAATCATCAGGTGTTAAACCAAAACCA 2501

```

```

RESULT 13
DDU20432
LOCUS      6115 bp      DNA      linear      INV 14-FEB-1995
DEFINITION Dictyostelium discoideum TagB (tagB) gene, complete cds.
ACCESSION  U20432
VERSION     U20432.1  GI:664839
KEYWORDS    Dictyostelium discoideum
SOURCE      Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 6115)
            Shaubsky,G., Kuspa,A. and Loomis,W.F.
            An MDR transporter/serine protease gene is required for prestalk
            specialization in Dictyostelium
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 6115)
            Shaubsky,G.
            Direct Submission
TITLE       Submitted (30-JAN-1995) Gad Shaubsky, Department of Biology 0322,
            University of California, San Diego, 9500 Gilman Dr., La Jolla, CA
            92093, USA
FEATURES             Location/Qualifiers
     source            1..6115
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="4"
                     /map="5.2 Mb from DIRS end"
                     /cell_type="prestalk"
                     /clone_lib="Lambda zap 10/45 (Shaulsky and Loomis,
                     Department of Biology 0322, Univ. of California, San
                     Diego)"
     misc_feature       1..6115
                     /phenotype="developmental arrest at tight aggregate stage,
                     PST-A cell development missing."
     gene              1..5886
                     /gene="tagB"
     mRNA              <1..5883

```

/gene="tagB"
 /product="tagB"
 <1..168
 /gene="tagB"
 /gene="tagB"
 /note="serine protease/MDR transporter homology, prestalk
 specific"
 /codon_start=1
 /product="tagB"
 /protein_id="AA62212.1"
 /db_xref="GI:664840"
 /translation="MKPQSPSPKIFLSSVILLIIFIGIKFELLEDNSNRNDKFN
 IIRFNINDDSIYKNOQOQFNSKIYSNEKKILLKNIIDITIKPSININNNNN
 NKLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
 DDQYKKNHYTVQPKDRINDETLKEFLIGDITLKEQPKSHVTHIPDSE
 LVPMTQSVLLSKKSWLWIGHEPNSKHLNHEKSGYVYIILSGNSLQIRW
 ENLNSLTSYNSKVKLTILNQKLLSVYCNDRSPSSSSSSCSLIGSEKIVYKWS
 EQESNYSERSEKLOTANRUSPTVIFGKDKLVNDRIDIPLRGGOILSIADTGLD
 SHGFFSDSKYPIFNQVNNHNRKVITYIHNDYVNGHGTGVCSSAGTPESSWA
 ISSPSGLATDAKIAFYDLSSGSEPTPPEDYSQMYKPLYDAGARVHGSWSVSGY
 YGVSDAGGIDAPLEYEPFSLRAAGNNELFALLAQATKNAITVGAETAHVNY
 VDALEYDFSDANFORPCLFDDKYNCTTAKCCSEVSNVKGLOLCCPASTKQNSD
 SPTPOFYNNENNGSSPKGTHDGRUKPDIVAPGEVITSARSNGENSTDOCGSL
 NNAOLMSTSGTSMATPLATAATLILROYLDVGEPTGESVEENKLLPTGSLKALMI
 PNAOLMSTSGTSMATPLATAATLILROYLDVGEPTGESVEENKLLPTGSLKALMI
 IGIGLGNOKATKEDSLSGLNKSICYFTYKPSSSSGSGGGGGTTPRIVATLVWTD
 PPSYSGAKFNLLNNDLILNSDDSIITIGNSGSLQOPAGKVAOPDTLNNVEGIIN
 PTKAMNYKFTAGTNPVIGPKQFSEVHENGQFWDSCQPCQVGPQPPCLITNGIG
 IQSGSLMTKTVQSCNENYNSLNTCDKFLSYNYIIVAGTGVLTILLLMW
 IKQYKESKEDSRFRDDGTGIFVRPKDKAKVTPDLYNLVSPFIILTIATACSL
 VATAASLOPPYIGNVNNIITTKISGEFSDFIILFIILAFIEFLTVNGSWTSGTVN
 EKMVRLQNVFRALIAQDMGFFQRNNSALLMNVILVDTPLMRSALTGILLSTATVC
 KLVSGLVFIPTISWKLFAFAVPILGLVTOIQSOFTKRLRLQLLFRNSKASQHTE
 SLTNHVVNTYCKQEKEMKYSDQMLNFIATKLIITNTEAGTKWLLIESLTFVIL
 YFSALVTOKFTGLMISFSLYIGVVPASSLFCVYVSYIOCLASATRVFMILKSA
 PRKTLTEERADRLAGLSGGGGGGGDDKDKKONIENGKDVLPNSIITPIDNVEN
 SNGKODPNNNNNIENLDYSEOLDGNDVADTVGLTKRELKQKKEKQKFCYGLV
 ISVATPTFLPSSVTELCRGEIEFKVNSFRYPTRPDVQVLHNI NKKEPAGKCYGLV
 ATIGENIYAIDNPQEDIIIEAKLAYAHEFIKVLTKQHTVVMVAHVAAVRCCKRI
 VARATCAKRLMLDEITAEIDPESEATIQSKIVLTQHTVVMVAHVAAVRCCKRI
 FVLKRGYLVBEETHDELMAKNGKYRMFSEDKDDTFLQNNNNNNNNNNNNPSS
 PTPNDQPTPPOEQEKNDQPPPPPOEQEQEQEQEQEQEQEQEQEQEQEQEQEQ
 QEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ
 NDBER"
 1300..2266
 /gene="tagB"
 /note="encodes serine protease homology"
 3166..4368
 /gene="tagB"
 /note="encodes MDR transporter homology transmembrane
 domain"
 4711..5238
 /gene="tagB"
 /note="encodes MDR transporter homology ATP binding
 domain"
 2376 a 844 c 926 g 1969 t
 BASE COUNT
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.77e-22 Length: 6115
 Score: 497.00 Matches: 162
 Percent Similarity: 40.48% Conservative: 72
 Best Local Similarity: 28.03% Mismatches: 162
 Query Match: 22.12% Indels: 182
 DB: 3 Gaps: 22
 US-09-985-689A-1 (1-434) x DD020432 (1-6115)
 Qy 19 LeuTyrGlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArg--- 37
 Db 1294 TTAAGAGGTAAGAGTCAGATATTGAGTATTGCTATACACTGCTTGTGTTAGTGGTAGGCATGTG 1353

38. -----AsnAspSer-----SerMetHisGluAlaPheArgGly 48
 1354 TTCTTTTCAGATTCAAAGTAGTCCCAATACCATTTAATCAAGTCAATGAAATCATACAGAAA 1413
 49 LysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAlaAsnAlaAspThrAsnGlyHis 68
 1414 GTTGTAACT-----TATATTACTTACCATGACAATGAAGATTATGTAATGTCAT 1464
 69 GlyThrHisValAlaGlySerValLeuGlyAsnGly-----Ser 81
 1465 GGTACACATGTTTGTGGCTCTGCAGCAGTACTCCAGAGATTCTTCATGGGGTATTTCAT 1524
 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 101
 1525 TCATTAGTGTCTTGCACACTGACCCAAAGATTGCAATTTATGAT---CTTCAATCIGGA 1581
 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
 1582 AGTTCTGAACCAACACCACCCGAGATTACAGTCAAATGTACAAACCATATATATGATGCA 1641
 122 GlyAlaArgIleHisThrAsnSerTrpGlyAla-----AlaValAsnGlyAla 137
 1642 GGTCAAGAGTACATGGTGTATCTTGGGGTCTCTGTATCTTTCGCAAGTTATTTATGGTGT 1701
 138 TyrThrThrSerSerArgAsnValAspAspTyrValArgLys---AsnAspMetThrIle 156
 1702 TATCCGATGATGCTGGTGTATGATGATGATCTCTATGATGATCCAGAAATTTCTCTATA 1761
 157 LeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAla 176
 1762 CTAAGAGCTGCTGGTAAT---AACGAGCTATTGTCATCTTTATGATGATGATGATGATG 1818
 177 LysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer----- 194
 1819 AAAATGCAATTTACAGTTGGTGTGAGCAACAGCTCATGTAATTTATGTCAGATGCA 1878
 195 -----TyrAlaAspAsnIle----- 199
 1879 TTGGAATATTGATGATTTCTCAGATAATGCTAATTTTCAAAGACCATGTTTATTCGATAAG 1938
 199 ----- 199
 1939 AAGTATTGTAATTATACAGCCGCTAAATGTTGCTCAGAGGTTTCAAATGTTTAAAGGTTTA 1998
 199 ----- 199
 1999 CAATTATGTTGTCAGCATCTAATAACAAATGTCATCGGATTCATTCAACACACACCT 2058
 200 -----AsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGly 214
 2059 CAATTTTAAATGAAAATAATATGGGATCATTCTCATCAAAAGGGTCCCAACACATGATGCT 2118
 215 ArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer--- 233
 2119 AGATTGAAACCTGATAGTGGACCTGGTGAATATATATACATCGGCAAGATCAATGATG 2178
 234 -----LeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyr 248
 2179 GAGAAATCAACAGACCAATGTGGTGGTGTCTTTA-----CCAAATGCAATGCTCTA 2232
 249 AlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
 2233 ATGTCATATCTGTATCAATGGAACACCATTTGGCAACAGCAGTGGCAACAGCAACAATTC 2292
 269 ArgGluHisPhe-----ValLysAsnArgGlyIle 278
 2293 AGACAATATTAGTTGATGTTATTTCCCAACTGGCGAATCTGTGCAAGAGATAAATACTA 2352
 279 ThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeu 298
 2353 CTACCAACTGATCATATTATAAAGCAATTAAATGATTAATGATTAATGCTCAATATTATAA 2412

```

QY 299 GlyTyr----- 300
Db 2413 ACTTATTGTTGGATGTCATCAAGTACAAATCCATCAACGCAATATTGAACAATTAAT 2472
QY 301 ---ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAla 319
Db 2473 GGCACAATTAATTAAGGTGGGTGCGTCACTAGAAATCAAT-----AATTGGTTA 2523
QY 320 TyrValAsnGluSerSer-----SerLeuSer 328
Db 2524 TATGTTAAATCATCAAAATCCACACCACCATCAAGATGGATTGGTATCGGTGGTTAGGT 2583
QY 329 ThrSerGlnLysAlaThr----- 334
Db 2584 AAAAATCAAAAGCAACTGAATGAAGAAAGATAGTTTATCAAGTGGTTAAATAAATCA 2643
QY 335 TyrSerPheThr-----AlaThrAlaGlyLys 343
Db 2644 TATTGTTTCACATATAAACCATCATCATCATCATCATCATCATCATCATCATCATCAT 2703
QY 344 Pro---LeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerVal 362
Db 2704 CCAAGAATTTGTCACATTAAGTTGGACAGATCCCGTCATATTTCTGGTGCAAAATTT 2763
QY 363 ThrLeuValAsnAspLeuLeu-----ValIleThr 373
Db 2764 AATTTAGTAATTAATTTAGATTTTATTATTATTAAATAGTATGATGATGATGATGATGAT 2823
QY 374 AlaProAsn-----GlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAsp 391
Db 2824 AUTGGTAATTCGTGGTGATCATTAACACACGACGAGGAAGGTAGCACAACTT----- 2874
QY 392 AsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThr 411
Db 2875 -----GATACATTAATTAATTTAGGGTATAATTAATTAATTAATTAATTAATTAATTA 2928
QY 412 TyrThrIleGluValGlnAlaTyrAsnValProValGlyProGlnThrPheSer 429
Db 2929 TATAAATTTACATCCGACGAACCAATGTACCAATTTGGACCAAAAGTTTTC 2982

RESULT 14
AR201152
LOCUS AR201152 1977 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 11 from patent US 6358726.
ACCESSION AR201152
VERSION AR201152.1 GI:20252040
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1977)
AUTHORS Takakura,H., Morishita,M., Shimojo,T., Asada,K. and Kato,I.
TITLE Thermostable protease
JOURNAL Patent: US 6358726-A 11 19-MAR-2002;
FEATURES
    Location/Qualifiers
        1..1977
            /organism="unknown"
BASE COUNT 453 a 659 c 554 g 311 t
ORIGIN

Alignment Scores:
Pred. No.: 9.25e-20 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 6 Gaps: 18

US-09-985-689A-1 (1-434) x AR201152 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla 27
Db 433 ATAGGGCCGATACCGCTGTGAACTCCCTCGGCTACGACGGAAGCGGTGTGGTGTGCC 492

```

```

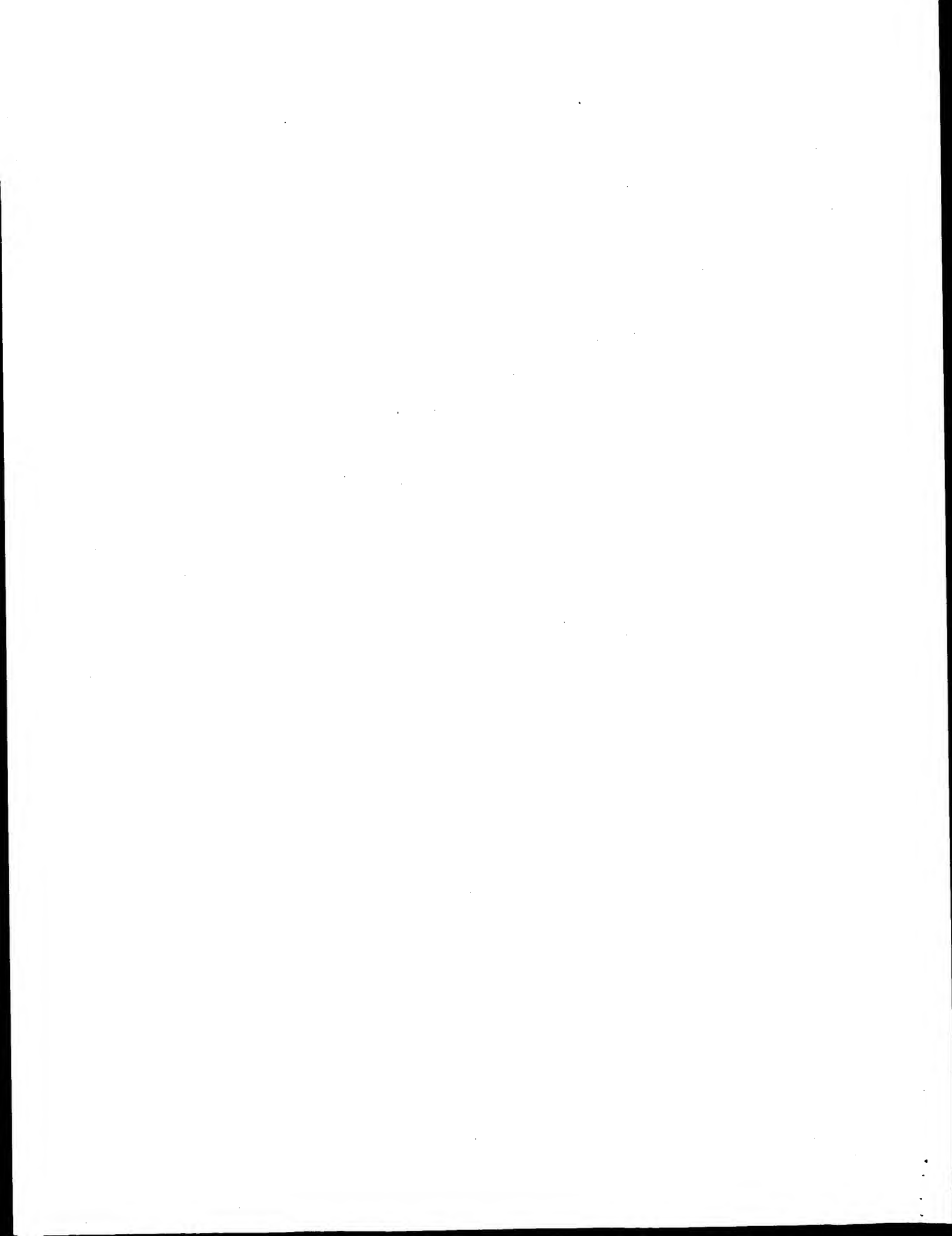
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisLluAlaPheArg 47
Db 493 ATCGTGCATACGGGTATAGACGCAAC-----CACCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GGCAGAGTCTAGGCTGTACGACGCGTCAACGGCAGGTGCGACCCCTACGATGACCCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACACGGAAACCCACGTTGCGGGTATCGTCCCGGAACCGCGCATCGTTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGGCTGCGCCCGCGGCAAGCTCGTCGGGTCAAGGTTCTCGGTGCGCGCGGTTCG 714
QY 102 GlyGlyLeuGlyClyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGCGGGTGTGACTGGTCTGCCAGAACAGCAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAspSer 129
Db 775 GGGATAAGGCTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGAGCAACCGATCC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
Db 835 CTCAGTCAGCGCGTCAACACGCGCTGGAGGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
Db 868 -----GGTATAGTCTGCGTCGCGCGCGCAACAGCGCGCGCAACACCTACACC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
Db 919 CTCGCTCACCCCGCGCGAGCAAGGTCATAACCGTCGCTGCCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 964 -----GTTGACAGCAACGACATCGGACGCTTCCACAGCGGGA 1005
QY 210 ProThrLysAspGlyArgLysProAspValMetAlaProGlyThrPheIleLeuSer 229
Db 1006 CCGACCGCGGACGGAAGGCTCAAGCGGAAGTCTGCGCCCGCGGCTTGACATCATAGCC 1065
QY 230 AlaArgSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAla 249
Db 1066 CCGCGCGCGCAGC-----GGAACGACGATGGGCACCCCGATAAAGCACTACTACACC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
Db 1117 AAGGCTCTGCAACACGATGGCACCCCGCAGCTTTCGGGGTGGCGGCGTCATCTCTC 1176
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
Db 1177 CAGGCCAC-----CCGAGTGGACCCCGCAAGGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
Db 1213 ACCGCTCTATCGAGACCGCGCATATCTCGCCCAAGGAGATAGCGGACATCGCCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
Db 1273 GGTGCG-----GGTAGGTGAACGCTCTACAGGCCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
Db 1312 GACGACTACGCCAAGCTCACTTCACCGGCTCCGTCGCGCAACAGGAAGGCCACCCAC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
Db 1372 ACCTTCGACCTACGCGCGCACCTTCTGTCACCGCCACCTCTACTGGGAC----- 1422

```



```
QY 202 -----ValAlaGlnPheSerSerArgGly 209
Db 1650 TATCAACCAATTCAACAGTTTATAGTAATTCATCCAGTTTATTTAGTGGTGTGGT 1719
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
Db 1720 CCAACATCTGATGTCGTTTGAACACAGATTATTAGCACCAGGTTCCACCAATTATTTC 1779
QY 230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHis----- 244
Db 1780 TCACGT---TCCTGGGTCCATCATCAACCAATT-----AATCATTTGTTCCACCAATTACA 1830
QY 245 -----AspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1831 TCAGGTATTGCAACATCGCTTTAATCGCTATGGAAGGTTCAAGTCAAGCAGCTGCTGTC 1890
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGlu-----HisPheValLysAsn 275
Db 1891 GCTACAAAGTCCCGCAGTTTATGTTGGTCAATTAATATAGAGATGGTTATTTTATTAATGGT 1950
QY 276 Arg-----GlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIle 290
Db 1951 AAAGTTAATTCATCTGTGGGATTCACCAACCATCAGCATCATTTGGTTAAAGCAACCTTAATC 2010
QY 291 AlaGlyAlaAla-----AspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrp 307
Db 2011 AATACTGCAAGTATTAAATGTTGATTCACATTAAGATAT-----TCACAAGGATTT 2061
QY 308 GlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSer----- 325
Db 2062 GGTAAATATTCAACTTTCAAGTTAATCACAACTACAATGCCAAACCAACCATCTTTAGAC 2121
QY 326 -----SerLeuSerThrSerGlnLysAlaThrTyr 335
Db 2122 ATTCATCATPAAATGAAAAGCTGATCCAAATAATTAATACTGCTGGAACCAATTCATAT 2181
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaPro 355
Db 2182 TGTTTCTCACTCGATAGTAAGCAGATATTGATATAACATTTGGTTTGGACAGATCCTGCC 2241
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAla--- 374
Db 2242 GGTTCACCAATTATCAACATTTACTTTGGTAAATAATTTGGATTAGCTTTACTTGCATTT 2301
QY 375 ProAsnGly-----ThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsn 392
Db 2302 GTTGATGGTGCACTTTCAATATATTCAGGAATAGTGAACCACTTTTCAAAAAATACAAGC 2361
QY 393 -----TrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSer 409
Db 2362 CAAGTAATTTTCGATCAATTAATAATGTTGAAGTCATTAAGATTAAGGATGCACCAATT 2421
QY 410 GlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyProGlnThrPheSer 429
Db 2422 GGTAGTTATGATGTTAAATAATTTGGTACAAATATTGTAATTCCAAAATCAATCCTATTCA 2481
QY 430 LeuAlaIle 432
Db 2482 GTTGTAAATT 2490
```

Search completed: July 28, 2003, 07:02:57
Job time : 5474.74 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model.

Run on: July 28, 2003, 03:42:25 ; Search time 100.115 Seconds
(without alignments)
1913.396 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EQAYNPVPGPTQFSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model
-Q=/cgn2_1/uspto_spool/US09985689/runat_25072003_143032_864/app_query.fasta_1.1166
-DB=Issued Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 -CGN_1_115=runat_25072003_143032_864 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*\n2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*\n3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*\n4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*\n5: /cgn2_6/ptodata/1/ina/PTOUS_COMB.seq:*\n6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	1923	4	US-09-509-814A-5
2	2242	99.8	1923	4	US-09-509-814A-7
3	2183	97.2	1920	4	US-09-509-814A-3
4	2125.5	94.6	3003	2	US-08-873-479-41
5	452.5	20.1	1977	3	US-08-894-818B-2
6	452.5	20.1	1977	3	US-09-445-472-11
7	416.5	18.5	1236	4	US-09-445-472-2
8	416.5	18.5	1566	3	US-08-894-818B-4
9	416.5	18.5	1962	3	US-08-894-818B-34
10	416.5	18.5	1962	4	US-09-445-472-15
11	403.5	18.0	1977	3	US-08-894-818B-6
12	351	15.6	2539	3	US-09-000-016-3

13	351	15.6	2539	4	US-09-514-340-3	Sequence 3, Appli
14	346	15.4	2809	3	US-09-000-016-1	Sequence 1, Appli
15	346	15.4	2809	4	US-09-514-340-1	Sequence 1, Appli
16	310.5	13.8	2532	1	US-07-671-376C-4	Sequence 4, Appli
17	307.5	13.7	1859	3	US-08-894-818B-15	Sequence 15, Appli
18	307	13.7	2835	1	US-08-750-532-2	Sequence 2, Appli
19	307	13.7	4765	1	US-08-750-532-8	Sequence 8, Appli
20	307	13.7	4765	3	US-08-894-818B-7	Sequence 7, Appli
21	307	13.7	4765	4	US-09-445-472-5	Sequence 5, Appli
22	296	13.2	898	1	US-08-750-532-7	Sequence 7, Appli
23	290.5	12.9	564	1	US-08-750-532-11	Sequence 11, Appli
24	290.5	12.9	564	3	US-08-894-818B-14	Sequence 14, Appli
25	276	12.3	1306	4	US-09-966-921A-1	Sequence 1, Appli
26	276	12.3	1330	4	US-09-966-921A-5	Sequence 5, Appli
27	256.5	11.4	1194	4	US-09-328-352-3407	Sequence 3407, Ap
28	256.5	11.4	10216	2	US-08-875-154-1	Sequence 1, Appli
29	252	11.2	1628	4	US-09-634-238-212	Sequence 212, App
30	251	11.2	840	1	US-08-434-255-7	Sequence 7, Appli
31	251	11.2	840	1	US-08-459-967-7	Sequence 7, Appli
32	251	11.2	840	1	US-08-460-327-7	Sequence 7, Appli
33	251	11.2	840	1	US-08-459-871-7	Sequence 7, Appli
34	251	11.2	840	3	US-09-024-532-1	Sequence 1, Appli
35	251	11.2	840	4	US-09-104-623A-1	Sequence 1, Appli
36	251	11.2	840	4	US-09-019-532-1	Sequence 1, Appli
37	251	11.2	840	4	US-09-417-359A-1	Sequence 1, Appli
38	251	11.2	1110	1	US-08-434-255-5	Sequence 5, Appli
39	251	11.2	1110	1	US-08-459-967-5	Sequence 5, Appli
40	251	11.2	1110	1	US-08-460-327-5	Sequence 5, Appli
41	251	11.2	1110	1	US-08-459-871-5	Sequence 5, Appli
42	251	11.2	1191	1	US-08-434-255-3	Sequence 3, Appli
43	251	11.2	1191	1	US-08-459-967-3	Sequence 3, Appli
44	251	11.2	1191	1	US-08-460-327-3	Sequence 3, Appli
45	251	11.2	1191	1	US-08-459-871-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-509-814A-5
Sequence 5, Application US/09509814A
Patent No. 6376227

GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO

APPLICANT: OKUDA, MITSUYOSHI

APPLICANT: SAEKI, KATSUHIISA

APPLICANT: KUBOTA, HIROMI

APPLICANT: HITOMI, JUN

APPLICANT: KAGEYAMA, YASUSHI

APPLICANT: SHIKATA, SHITSUW

APPLICANT: NOMURA, MASAFUMI

TITLE OF INVENTION: ALKALINE PROTEASE

FILE REFERENCE: 0327-0832-0PCT

CURRENT APPLICATION NUMBER: US/09/509,814A

PRIOR FILING DATE: 2000-04-06

PRIOR APPLICATION NUMBER: PCT/Jp98/04528

PRIOR FILING DATE: 1998-10-07

PRIOR APPLICATION NUMBER: JP 9-274570

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Patentin version 3.0

SEQ ID NO 5

LENGTH: 1923

TYPE: DNA

ORGANISM: Bacillus sp.

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1923)

US-09-509-814A-5

Alignment Scores:

Pred. No.: 1.94e-221

Score: 2247.00

Percent Similarity: 100.00%

Length:

Matches: 1923

Conservative: 434

Handwritten: Nucleic acid

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1 (1-434) x US-09-509-814A-5 (1-1923)

QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 619 AATGATGTTGGCGTGGAAATTTGCAAGCGGATGGCTCAGACGACGTACCGGTTGTAT 678
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGACAGGACAGATCGTACGGTTCGCCATACAGGGCTTGATACAGGTGCGAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 739 TCGATGATGAAGCTTCGCGGGGAAATTTACTGCATTTATGATGGGACGCGAGAT 798
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 799 AATGCCAATGATACGAATGGTTCATGTCGCGATGGCTCCGATTTAGGAACGGC 858
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCAATCTATCATGGATAG 918
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 919 GGTGGGGGACTTGGAGGACTACCTTGAATCTGCAAACTTATTCAGCCAGGATACAGT 978
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
DB 979 GCTGGTGCAGAAATTCATACAACTCCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1039 GATTCAGAAATGTGGATGACTATGTCGCAAAATGATATGACGATCCTTTTCGGCTGCC 1098
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 1099 GGAATGAAGGACCGAAGCGGGAACCATCAGTGCACCGCAGCAGCTAAAAATGCAATA 1158
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 1159 ACAGTCGGAGCTACGGAACCTCCGCCAAAGCTTTTGGTCTTATGGCGCAATATCAAC 1218
QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 1219 CATGTGGCAGATCTCTTCACGTGGACCGCAGCAAGATGGACGATCAACACCGATGTC 1278
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 1279 ATGGCACCGGAACGTTTACTATCATCATCAGCAAGATCTTCTTTCGACCGGATTCCTCTC 1338
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
DB 1339 TGGCGCAACCATGACAGTAATAATGATATGATGGTGGAAAGTCCATGGCTACACCGATC 1398
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 1399 GTTGTGGAAACGTGGCAGCAGCTTCGTGAGCATTTTGTGAAACACAGAGGATCACACCA 1458
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 1459 AAGCCTTCTCTATTAAGCGGCACTCATTCGCCGCTGCAGCTGATCGGCTTGGCTAC 1518
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 1519 CCGAAGCGGTAAACAGGATGGGACGAGTGACATGGATAATCCTCGAAGCTTCCCTAT 1578
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 1579 GTGACAGGCTCCAGTCTCTATCCACCAGGCCAAAGGACGCTACTCGTTTACTGCTACT 1638

RESULT 2
US-09-509-814A-7
; Sequence 7, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUM
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-7

Alignment Scores:
Pred. No.: 6,35e-221 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 1
Query Match: 99.78% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1 (1-434) x US-09-509-814A-7 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 619 AATGATGTTGGCGTGGAAATTTGCAAGCGGATGGCTCAGACGACGTACCGGTTGTAT 678
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 679 GGACAGGACAGATCGTACGGTTCGCCATACAGGGCTTGATACAGGTGCGAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 739 TCGATGATGAAGCTTCGCGGGGAAATTTACTGCATTTATGATGGGACGCGAGAT 798
```


Db 739 TCGATGATGAAGCGCTTCGCGGGAATAATTACTGATATATGATGGGACGACGAAAT 798
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 858
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCCACATAAATAAGGAATGCGCCTCAGCGCAATCTAGTCTTCCATCTATCATGGATAGC 918
QY 101 GlyGlyGlyGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 919 GGTGGGGGACTTGGAGGACTTACCTTCGAATCTGCAAACTTATTCAGCCCAAGCATACAGT 978
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAlaAsnGlyAlaTyrThr 140
Db 979 GCTGGTGCCAGATTCATACAACTCTGGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1039 GATTCAGAAATCTGATGACTATGTCGCAAAATGATATGACGATCCTTTTCGCTGCC 1098
QY 161 GlyAsnGluClyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GCGAATGAAGACCGACCGGAGCACTCAGTGCACCGACGACGCTAAATAATGCAATA 1158
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTCGGAGCTACGCAAACTCTCGCCCAAGCTTTGGGTCTTATCGGCACATATCAAC 1218
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCACAGTCTCTTCAGTGGACCGGACGAAAGGATGGACGATCAACCGGATGTC 1278
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1279 ATGGCACCGGAACCTTCATACATATCAGCAAGATCTTCTTGCACCGGATCCCTCTTC 1338
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1339 TGGCGGAACCATGACAGTAATAATGATACATGCTGGTGAACGTCATGCTACACCGATC 1398
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1399 GTTCTGGAAACGTGGCACACCTTCGTGAGACATTTTGTGAAACAGAGGATCACACA 1458
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1459 AAGCTTCTCTATTAAGCGGCACATGATTCGCGGTGCAGCTGACATCGGCTTCGGCTAC 1518
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1519 CCGAACCGTAAACCAAGATGGGACGAGTACATGGATAATCCCTGAAGCTTCCTAT 1578
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1579 GTGAACGAGTCCAGTCTCTATCCACAGCCAAAGGAGGAGTACTCGTTTACTGCTACT 1638
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1639 GCCGGCAAGCTTTGAATAATCTCCCTGATGTGTCATGTCGCTGTCGAGCACAACTGCT 1698
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1699 TCCGTAACGCTTGTCAATGATCGGACCTTGTCTATACCGCTCCAAATGGCACACAGTAT 1758
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
Db 1759 GTAGGAATGACCTTTTCTTCGCCATACAATGATAACTGGGATGSCCGCAATAACGTAGAA 1818
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1819 AATGTATTTATTAAATGCACACAAAGCGGACGATACATAATTGAAGTACAGGCTTATAAC 1878

QY 421 ValProValcylProGlnThrPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGTTGGACCACAGAACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 3

US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1920)
US-09-509-814A-3

Alignment Scores:

Pred. No.: 7,45e-215 Length: 1920
Score: 2183.00 Matches: 418
Percent Similarity: 99.31% Conservativeness: 13
Best Local Similarity: 96.31% Mismatches: 3
Query Match: 97.15% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1 (1-434) x US-09-509-814A-3 (1-1920)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATCATGTGCCAGAGGATTTGTCAAAGCGGATGTGCCACAGACGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAGGCGAGATTTGCGCAGTTGCCGATGTGGATGTGGATACAGGAAGAACGACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAAGCCCTTCGCGGTAAATAACAGCACTATATGACCTGGGTCGGACGAAAT 795
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGGCAATGATACGAACGTCATGCTACCATGTGCCAGTTTCGGATTAGGAATAGGC 855
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACCAATAAAGGAATGACCTCAAGCGAAATCTGGTTTTTCAATCCATCATGGATAGC 915
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGTGGGCTTGGAGGCTTGCCTTCCAACTCTCAAACTTATTCAGCCCAAGCATTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140

Tue Jul 29 14:12:39 2003

us-09-985-689a-1.p2n.rni

ADDRESSEE: No. 58917010 No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agriis, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:

INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3003 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 1,24e-208 Length: 3003
Score: 2125.50 Matches: 406
Percent Similarity: 97.93% Conservative: 19
Best Local Similarity: 93.55% Mismatches: 8
Query Match: 94.59% Indels: 1
DB: 2 Gaps: 1

US-09-985-689a-1 (1-434) x US-08-873-479-41 (1-3003)

976 GCAGGTGCCAGAAATTCATAAACTCCTGGGGGCGAGCGGTGAATGGGGCCTACAGACA 1035
141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
1036 GATTCACAGAAATGGATGACTATGATAGAGAAATGATATGACGATTTCTTTTCGGCGCT 1095
161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
1096 GGGATGAAGCCGACGCGGTACCATCAGTGCACCTGGTACGGCTAAAAAGCGCAT 1155
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
1156 ACAGTCGGCGCAACCAACCTGGCTCGCAAGCTTCGGTTCCTATGAGATATATTAAC 1215
201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
1216 CACGTGACAGATCTCTTCCTCCGTGGCGGACACAAAGATGGCGCAATCAAGCCTGATGTC 1275
221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
1276 ATGGCGCCAGGACATACATATTTATCAGCAAGATCTCTCTGCACCCGATTCCTCTTC 1335
241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetClyGlyThrSerMetAlaThrProIle 260
1336 TGGCGAATCATGACAGCAAAATATGCTATATGGTGGACGCTCATGGCAACACCGATT 1395
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
1396 GTTGGCGGAATGTTGCACAGCTCCGTGAGCATTTGTGAAATATAGAGGAATCACTCCT 1455
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
1456 AAGCCTTCCTATTTGAAGACAGCTTTGATTCAGGCTGCTGATGTTGGATTGGGTTAT 1515
301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
1516 CCGAACGGAAACCAAGGATGGGCGGAGTGCACCTGGATAAATCGTTGAACGTTGCCAT 1575
321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
1576 GTGAACGAATCCAGTCCCTATCAACTAGCCAAAGGACATATACCTTTACTGCAAG 1635
341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
1636 GCGGGCAAGCAATGAAATCTCCCTGATGTCGGATGCCCTGCAAGCACTACTGCT 1695
361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
1696 TCTGTAACTTCAATGATTTGGATTTGGTCTATACAGCACCACCAAGGACAGATAT 1755
381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
1756 GTCGGGATGACTTCTCAGACCAATTTGACAAATTAACCTGGATGGCGGCAATTAACGTAGAA 1815
401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
1816 AATGATTTATTAATTCGCCCAAGTGAACATATACCATTTGAGGTGCAAGCATATAAT 1875
421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
1876 GTGCCGTTGGACCAAAACTTCCTGTTGGCAATTTGGAAC 1917

RESULT 4

US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lynne, Christanson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:

```

QY 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAATGAGGACAGGAGTAGCGGTACATCAGTCAGCAGGAGCAACAAAAATGCGATT 2006
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGCAACCGAAACCTAGTCCAGCTTCGGATCTTATGCGGATATTAAC 2066
QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATTCCTCAGGAGGCTCTACTAGATGGACGTATTAAAGCGGACGTC 2126
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 2127 ATGCAACAGGTACGTATATCTCTGCTAGATCATCATTTAGCTCCAGATTCCTCATTC 2186
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 2187 TGGCAACCATGATAGTAATATGCTACATATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2246
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 2247 GTAGCAGGTAATGTGCACAAATTAAGGAGCATTTTGIGAAAAATAGAGGGGTAACTCCT 2306
QY 281 LysProSerLeuLysAlaLeuLeuAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCCTTCCCTTTTAAAGCTGCTTTAATGTCAGGCTGCTGGGATGTTGGACTTGGCTTT 2366
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATGGTACCAAGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTGCGATT 2426
QY 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 2427 GTGAATGAAAGAGGACCTTTTATCAACAAGTCANAAAGCACATATTCGTTACGGCTCAA 2486
QY 341 AlaGlyLysProLeuLysIleSerLeuValThrSerAspAlaProAlaSerThrAla 360
Db 2487 GCTGGTAAACCTTTAAATAATATACATCTGTTGTCAGATGACGAGGAGTACGACGCA 2546
QY 361 SerValThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 2547 TCACCTAACCTTAGTGATGATTTAGACTTAGTAAATCAGTCACCAAAATGGAACCTAAATC 2606
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 2607 GTCGGAATGACTTTACAGCAGCGTATGATAACAATGGGATGGCAGAAACACGTCGAA 2666
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGTTTATCAATGCTCCTCAAGCGGAACGTATACAGTCGAAGTGCAGGCTTACAAT 2726
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAGTAAGTCCGCAACCTTTCTTTAGCGATTGTACAT 2768

```

RESULT 5

US-08-894-818B-2

; Sequence 2, Application US/08894818B

; Patent No. 6261822

; GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru

; APPLICANT: MORISHITA, Mio

; APPLICANT: YAMAMOTO, Katsuhiko

; APPLICANT: MITTA, Masanori

; APPLICANT: ASADA, Kiyozo

; APPLICANT: TSUNASAWA, Susumu

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES

; NUMBER OF SEQUENCES: 42

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Browdy and Neimark

```

; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA: JP 323285/1995
; APPLICATION NUMBER: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-2

```

Alignment Scores:

```

Pred. No.: 9,03e-37 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 3 Gaps: 18

```

US-09-985-689A-1 (1-434) x US-08-894-818B-2 (1-1977)

```

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla 27
Db 433 ATAGGGGCGGATACCGTCTGGAACCTCCCTCGCTAGCAGCGAAGCGGTGGTGGTGGC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCGTCGATACGGGTATAGACCGCGAAC-----CACCCCGCATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GGCAAGGTCATAGGTCGTGACGACCGCTCAACGCGAGGTCCACCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACAGGAACCCCGGTCGCGGTATCGTTCGGGTCAGGTTCTCGGTGCCACCGGTTCG 714
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGCGGTGCCCCCGGCGGCAAGCTCGTCGGGTCACAGGTTCTCGGTGCCACCGGTTCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGTCTGCCAACAAGACAGGATAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAAGGGTCAATCAACCTCTCCCTCGGCTCTCTCCAGAGCTCCGACGGAACCGACTCC 834

```

```

; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; US-09-445-472-11

Alignment Scores:
Pred. No.: 9,03e-37 Length: 1977
Score: 452.50 Matches: 138
Percent Similarity: 44.66% Conservative: 67
Best Local Similarity: 30.07% Mismatches: 153
Query Match: 20.14% Indels: 101
DB: 4 Gaps: 18

US-09-985-689A-1 (1-434) x US-09-445-472-11 (1-1977)

QY 8 VallysAlaAspValAlaGlnSerSerTyrGlyGlnGlyGlnIleValala 27
Db 433 ATAGGGCGGATACCGTGTGGAACCTCCCTCGGTAGCAGGAGCGGTGGTGGTGC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATGTCGTGATACGGGTATAGACGGCAAC-----CACCCCGCATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GGCAAGGTCTATAGGCTGTGACGACGCGGTACAGCGGAGGTGCGACCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACACGGAACCCACGCTTCGGGTATCGTTCGGGATCGTTCGGGCGTCAAGGTCTCGGTGCGGACGGTTCG 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGCGGTGCGCCCGCGGCGGAGAGCTCGTCGGCGTCAAGGTCTCGGTGCGGACGGTTCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 CGAAGCGTCTCCACCATCATCGCGGGTGTGACTGGTGGTCCAGCAACAGGACAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAAGGTCTATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGAGCGAACCGACTCC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
Db 835 CTCAGTCAGCGCTCAACACGCGCTGGAGCGC-----HisThrAsnSer 129
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
Db 868 -----GGTATAGTGTCTCGTTCGGCGCGGCAACAGCGGCGGCAACCGTTCACAC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGlyAsnLeuArg 189
Db 919 GTCGGCTACCCCGCGCGGAGGAGTATACCGCTCGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 964 -----GTTGACAGCAACGACAACTATCGGCGAGCTTCTCCAGCAGGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229

; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; US-09-445-472-11

130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
Db 835 CTCAGTCAGCGCTCAACACGCGCTGGAGCGC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
Db 868 -----GGTATAGTGTCTCGTTCGGCGCGGCAACAGCGGCGGCAACCGTTCACAC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
Db 919 GTCGGCTACCCCGCGCGGAGGAGTATACCGCTCGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 964 -----GTTGACAGCAACGACAACTATCGGCGAGCTTCTCCAGCAGGGGA 1005
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229

; Sequence 11, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
```

```
Db 1006 CCGACCGGCGGCGGAGGCTCAAGCGGAGTCTGCGCCCGCGCTTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAla 249
Db 1066 CCGCGCGGCGGCGGAGGCTCAAGCGGAGTCTGCGCCCGCGCTTGACATCATAGCC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
Db 1117 AAGGCGCTCGGAACAGCATGCGCACCGCGGAGTCTGCGCGGCTTGCGCGCTCATCCTC 1176
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
Db 1177 CAGGCGGCGGAGGCTCAAGCGGAGTCTGCGCGGAGTCTGCGCGGCTTGCGCGCTCATCCTC 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
Db 1213 ACCGCGCTCATCGAGCGGCGGAGTCTGCGCGGAGTCTGCGCGGAGTCTGCGCGGCTCATCCTC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
Db 1273 GGTGCG-----GGTAGGTTGAACGTTCTACAGGCCCATCAAGTAC 1311
QY 317 ---AsnValAlaThrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
Db 1312 GAGGACTACGCAAGCTCACCTTCACCGGCTCGCGCGGAGTCTGCGCGGAGTCTGCGCGGCTCATCCTC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValThrSerAspAlaPro 355
Db 1372 ACCTTCGAGCTCAGCGGCGGCGGAGTCTGCGCGGAGTCTGCGCGGAGTCTGCGCGGCTCATCCTC 1422
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuValIleThrAlaPro 375
Db 1423 -----ACGGGCTCGAGGAGTCTGCGCGGAGTCTGCGCGGAGTCTGCGCGGCTCATCCTC 1461
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGly 395
Db 1462 AACGGGAACGAG---GTTGACTACTCTACACCGGCTACTAC----- 1500
QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGlu 415
Db 1501 -----GGCTTCGAGAGGTCGCGTACTACACCGGAGTCTGCGCGGAGTCTGCGCGGCTCATCCTC 1554
QY 416 ValGlnAlaThrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1555 GTGCTCAGCTACAG-----GGCGGCGGAGTCTGCGCGGAGTCTGCGCGGCTCATCCTC 1602
```

RESULT 7

```
US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2
```

```
Alignment Scores:
Pred. No.: 2.18e-33 Length: 1236
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 4 Gaps: 20
```

US-09-985-689A-1 (1-434) x US-09-445-472-2 (1-1236)

```
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGI 24
Db 24 GTCGCAAGCTCAAGCTTATGGCAACTTACGTTTGGAACTTGGGATATGATGGTCTGGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACAATAGGAATAATTGACACTGCAATGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAAGGAAGATA-----ATTGGTGGGTAGATTTTGTCAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGGAGTTATCCATACCATGACCATGACATGGAACATCATGTAGCTTCAATAGCAGCTGG 230
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAAGGAATGGCTCCAGGAGCTAAGCTGCGGG 290
QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuG1 112
Db 291 AATTAGGTTCTAGTGCCGATGTTCTGGAACATATCTACTATAATTAAGGAGTGA 350
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGGCGGTGATAACAAAGATAAGTACGGAATTAAGGTCTAATTAATCTTCTTGGTTC 410
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 411 AAGCCAGACTCAGATGGTACTGACGCTCTAAGTCAAGCTGCTTAATCAGCTGGATGC 470
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTTGTTGGTTCGCG 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 495 TGGAAACAGTGACCTTCTCAAGCAGAGGGCCAACTGCAGAGCGGCTTAAGCTGAGGT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGAGCC-----GTTCAACAGTATGA 581
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATAACAAGCTTCTCAAGCAGAGGGCCAACTGCAGAGCGGCTTAAGCTGAGGT 641
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerph 240
Db 642 TGTGCTCCAGAAACTGGATAATTGCTCCAGAGCAAGT-----GGAAGTACGAT 692
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI1 260
Db 693 GGTCAACCAATTAATGACTATTACACACAGCTCTCTGGAGCATCAATGGCACTCTCA 752
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGTATTGAGCGCCCTCTTCTCTCAA-----GCACACCC 791
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGACAAGTAAAAACAGCCCTCATAGAACTGTGTATATCGTAAAGCC 851
```

```

294  QY -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
      |||||
852  Db AGATGAATAGCGCATATACCTACGGTGCA-----GGTAGGCT 890
      |||||
310  QY lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
      |||||
891  Db TAATGATACAGGCTATAAAC-----TAGGATAACTATGCAAGCTAGTGTTCACCTGG 944
      |||||
330  QY rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProle 345
      |||||
945  Db ATATGTGGTCCCAAGAGCCCAACTCCAGTTCGTTATTACGGAGCTTCGTCCT 1004
      |||||
345  QY uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
      |||||
1005 Db AACTGCCACATTTACTGGGACAATGCCAAT-----1035
365  QY lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
      |||||
1036 Db -AGCGACCTTGATCTTTACCTCTACGATCCCAATGGAAACCCAG---GTTGACTACTCTTA 1091
      |||||
385  QY eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
      |||||
1092 Db CACGCGCTACTAT-----GGATTGCAAAAGGTTGGTTATTATTA 1127
405  QY nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
      |||||
1128 Db CAACCCCACTGATGGAACATGGACAATTAAGGTTGTAAGCTACAGC-----GGAAG 1178
425  QY oGlnThrPheSerLeuAlaIleValAsn 434
1179 Db TGCAACTATCAAGTAGATGTTGGTAAAGT 1206

RESULT 8
US-08-894-818B-4
; Sequence 4, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/Jp96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1566 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; FEATURE:
; OTHER INFORMATION: /note= N at position 1283 is G or T.
US-08-894-818B-4

Alignment Scores:
Pred. No.: 3,17e-33 Length: 1566
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 3 Gaps: 20

US-09-985-689A-1 (1-434) x US-08-894-818B-4 (1-1566)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGl 24
Db 24 GTCTGCAGCTCAAGTTATGGCAACTTACCTTTGGAACTTGGGATATGATGCTTCTGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGl 44
Db 84 CACAATAGGAATAAATGACACTGGAATGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAAGGAAAGTA-----ATTGGGTGGTAGATTGTGCAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 171 TAGGATTAATCATACGATGACCATGGACATGGAACATCATGACTTCAATAGCAGCTGG 230
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGCAGCTAAGCTGGCGG 290
QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuGl 112
Db 291 AATTAAGGTTCTAGGTCCCGATGTTCTTGGAAAGCAATATCTACTAATAAGGAGTTGA 350
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCCGTTGATACAAAGATAAGTACGGAAATAGGTCATTAACTCTTCTCTTGGTTC 410
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 411 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAAGCTGCTTAATGAGCTGGGATGC 470
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTTGTGTGTTGCTGCCGC 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 495 TGGAAACAGTGGACCTAACCAAGTATACAATCGGTTCTCCAGCAGCTGCAAGCAAAAGTTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGCAAGTATGA 581
QY 200 nHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATTAACAAGCTTCTCAACGACAGGAGGCCCAACTGCAGACGCGAGCTTAAGCCTCAGGT 641
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240
```

[illegible]

RESULT 9

```

US-08-894-818B-34
: Sequence 34, Application US/08894818B
: Patent No. 6261822
: GENERAL INFORMATION:
: APPLICANT: TAKAKURA, Hikaru
: APPLICANT: MORISHITA, Mio
: APPLICANT: YAMAMOTO, Katsuhiko
: APPLICANT: MITTA, Masanori
: APPLICANT: ASADA, Kiyozo
: APPLICANT: TSUNASAWA, Susumu
: APPLICANT: KATO, Ikuoshin
: TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Browdy and Neimark
: STREET: 419 Seventh Street N.W., Ste. 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: United States of America
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30

```



```

; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
;
US-09-445-472-15

Alignment Scores:
  pred. No.: 4,52e-33      Length: 1962
  Score: 416.50           Matches: 141
  Percent Similarity: 42.55%      Conservative: 59
  Best Local Similarity: 30.00%    Mismatches: 148
  Query Match: 18.54%            Indels: 122
  DB: 4                      Gaps: 20

US-09-985-689A-1 (1-434) x US-09-445-472-15 (1-1962)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAAT 479
QY 24 nileValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGGAAGAAGTA-----ATTGGGTGGGTAGATTGTCATATGG 566
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAATCATGAGTCTTCAATAGCAGCTGG 626
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGCAGCAGCAAGTAATGGCAAGTACAAAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 686
QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAAGTTCTAGGTGCCGATGGTCTTGGAAACATATCTACTATATTAAGGAGGTGA 746
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAl 132
Db 747 GTGGCCCGTTGATACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTGTGTTTC 806
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGCAGCTGGGATCC 866
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTGTGTGGTGGCGC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 891 TGGAAACAGTGGACCTTAACAAGTATACAAATCGGTCTCCAGCAGCTGCAAGCAAGATTAT 950
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGACAAGTATCA 977
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATATAACAAGCTTCTCAAGCAGAGGGCCCAACTGCAGACGCGCGCTTAAGCCCTCAGGT 1037
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGTCTCCAGAACTGGGATAATTGCTGCCAGAGCAAGT-----GAACTAGCAT 1088
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI1 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCAACTCCTCA 1148
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGSPATTGCAGCCCTCTGTCTCCAA-----GCACACCC 1187
QY 280 oLys-----ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGGACTCCACACAAAGTAAACAGCCCTCATAGAACTGCTGATATCCTAAAGCC 1247
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 1248 AGATGAATAGCCGATATAGCTACGGTGCA-----GGTAGGGT 1286
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 1287 TAATGCATACAGGCTATAAC-----TACGATAACTATGCAAGAGTAGTGTCTACTGG 1340
QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAAACAAAGCAGCCAACTCACCAGTTCTGTTATTAGCGAGCTTCGTCTGT 1400
QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTAATCTAGTGGGCAATGCCAAT----- 1431
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTTACCTCTACGATCCCAATGGAACCCAG---GTTGACTACTCTTA 1487
QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCCTACTAT-----GGATTCGAAAAGGTTGGTTATTA 1523
QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGAACATGGCAATTAAGGTTGTAGCTACAGC-----GGAAG 1574
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGGTAAGT 1602

RESULT 10
US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06

```



```

QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCCCTCA 1148
QY 260 evalAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeThrPr 280
Db 1149 CGTAGCTGGTATTCAGCCCTTCTGCTCAA-----GCACACCC 1187
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGAGCTCCAGCACAAGTAAACACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 1248 AGATGAATAAGCCGATATAGCTACGGTGCA-----GGTAGGGT 1286
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 1287 TAATGCATACAGGCTATAAC-----TAGGATAACTATGCAAGAGTAGTCTTCACTGG 1340
QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAAGCGCAAACTCACCAGTTCGTTATTAGCGGAGCTTCGTTGCT 1400
QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTAATGCTGGGACCAATGCCAAT----- 1431
QY 365 lAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTACCTCTACGATCCCAATGGAACACAG---GTTGACTACTCTTA 1487
QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1488 CACCGCCTACTAT-----GGATTGCAAAAGGTGTTGTTATTA 1523
QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAAACATGGACAATTAAGGTGTGAAGTACAGC-----GGAAG 1574
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGTGGTAAGT 1602

```

RESULT 11

```

US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894.818B

```

```

; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-6

```

```

Alignment Scores:
Pred. No.: 9,95e-32 Length: 1977
Score: 403.50 Matches: 134
Percent Similarity: 43.07% Conservative: 68
Best Local Similarity: 28.57% Mismatches: 147
Query Match: 17.96% Indels: 120
DB: 3 Gaps: 20

```

US-09-985-689A-1 (1-434) x US-08-894-818B-6 (1-1977)

```

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTCGACCTCAAGTTATGCGCAACTTACGTTTGGAACTTGGGATATGATGGTCTGGAAT 479
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCCAAGAAAGTA-----ATTGGGTGGTGTAGATTTCGTCAATGG 566
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTATATCCATACGATGACCATGGACATGGAATCATGAGTTCATACACAGCTGG 626
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGGAAATGGCTCCAGGAGCTAAGCTGGCGGG 686
QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAGGTCTTAGTGCCGATGGTCTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle----- 125
Db 747 GTGGGCGGTTGATAACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTGGTTC 806
QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCCGACGACCGACTCCCTCAGTCAGCGGTCACACAGCTGGAGC 866
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla1 160
Db 867 C-----GGTATAGTAGTCTCGCTGCCGCC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla1 180
Db 891 CGGCAACAGCGGCGGCAACACCTACACCGTGGGTCTACCCCGCGCGGCAAGGTTCAT 950

```



```
Db 1286 GCGCCGCGCGAGCGCGCGCGTCTGTCACCATGACGCTGGGC-----GSCATG 1333
QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetThr 155
Db 1334 GACACCGGAGACCGCGCGCGCGCGTGTGCGACAGCGTCTCCCGCGAGAGAGCGC 1393
QY 156 IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCCTGTGCGCATCGCGCGCGCGCAACGAGGCGCGCGAG-----TCGATCGGTTCGCC 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGAGCGCGCGCGTACCGTGGCGCG----- 1480
QY 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTCGACGACAGGACAGCTCGCGACTTCCTCCACGCGCGCGCGCTCGGC 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGGCGCATCAAGCGCGAGCTCACCGCTCCCGCGTGGACATCACGGCGCGCTCGCGCG 1594
QY 233 SerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GACGGCAACGACATCGCGCGAGGAGTCTGTCAGGACCGCGCGGTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCACGTCGATCGCGACCGCGCGCGTGGCGCGCGCGCGTCTCTGAGCAGCAG--- 1711
QY 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCGCGACTGGACCTCCGCGCGAAGTCAAGGGCGCG 1747
QY 290 IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg 309
Db 1748 CTCACCGGTCCACCAAG---GGCGCAAGTACACCCCGTTCGACGAGGGTTCGGGCGCG 1804
QY 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer--- 328
Db 1805 ATCCAGCGCGACAGCGGTCCAGCAGACGCGTGTATCGCGCGACCCCGTTCGACGAGGGTTCGGGCGCG 1864
QY 329 -----ThrSerGlnLysAlaThrTyrSer 336
Db 1865 GCGTCCACGAGTGGCGCGCACACCGACGACGCGCGCGCGCGTCCACCAAGCAGCTGACCTACCGC 1924
QY 337 PheThrAlaThrAlaGlyLysProLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGCACCCAGCAGCTCAGCTGAGCTGAGCTGACCTGACCGCCACCGACCCCAAGGGC 1984
QY 354 -----AlaProAlaSer 357
Db 1985 AAGCGCGCGCGCGCGGTCTTTCAGCTGGCGCGCGCGCGCGTCCCGCGCGCGC 2044
QY 358 ThrThrAlaSerValThrLeuValAsnAspLeuAspLeu----- 370
Db 2045 GGCAGCGCTCGCTCGACATACCGCGCGCGCGCGCGCGTTCGGCGCGCGTGGAGCGCGCG 2104
QY 371 -----ValIleThrAlaProAsnGlyThrGluTyrVal 381
Db 2105 TACTCGCGCTAGTGTGCGCGCGCGCGCGCGCGCGTCCAGCGGTC 2146
```

RESULT 13

```
US-09-514-340-3
; Sequence 3, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIVATIVES
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
```

```
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-514-340-3
```

```
Alignment Scores:
Pred. No.: 3,728-26 Length: 2539
Score: 351.00 Matches: 131
Percent Similarity: 42.63% Conservative: 54
Best Local Similarity: 30.18% Mismatches: 159
Query Match: 15.63% Indels: 90
DB: 4 Gaps: 16
```

US-09-985-689A-1 (1-434) x US-09-514-340-3 (1-2539)

```
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACAGCTCGTGGCGCAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGTCTGACACCGCGTGTGACACGAGC----- 1054
QY 42 MethisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCGGACCTGAAGGCGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCGCGCGACAAAGTGGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAGGCGCAAGTACAGGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1231
```

QY 98 MetAspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACACACCTCCGT-----TTGGCGACAGCTCCGCGATCTCCCGCGATGGAGTGG 1285
QY 118 AlaThrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
Db 1286 GCGCGCGCGAGCGCGCGAGCTCGTACCATGAGCTGGG-----GGCATG 1333
QY 138 TyrThrThrAspSerArgAsnValAspAspThrValArgLys-----AsnAspMetThr 155
Db 1334 GACACACCGGAGACCCCGCTGGAGCGCGGCGTGCAGACAGCTGTCCCGCGAGAGGCG 1393
QY 156 IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCTGTTCGCCATCGCGCGCGGCAACAGAGGCGCGGAG-----TCGATCGTTCGCC 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGGAGCGCGCTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCG 1480
QY 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLys 212
Db 1481 -----GTGCGACGACAGGACAGCTCGCGGCTTCTCTCCCGCGCGCGCTCGCG 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGCGCGCATCAAGCGCGAGCTACCGCTCCCGGCTGCACATCAGCGCGCGCTCGCG 1594
QY 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaThrMetGly 252
Db 1595 GAGGCGACGACATCGCGCGAGGCTCGGTGAGGAGCGCGCGCTCGCGTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCAGCTCGATGCGCGCGCGCGCGCTCGCGCGCGCGCGCTCTCTGAGCAGCAG--- 1711
QY 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCGCGACTCGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1747
QY 290 IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArg 309
Db 1748 CTCACCGCTCCACCAAG---GGCGGCAAGTACACCCGCTTCGAGCAGGCTTCGCGCGCG 1804
QY 310 ValThrLeuAspLysSerLeuAsnValAlaThrValAsnGluSerSerSerLeuSer--- 328
Db 1805 ATCCAGCGCGCAAGCGCTCCAGCAGACCGGTATCGCGCGCGCGCGCTCTCGGTGAGCTTC 1864
QY 329 -----ThrSerGlnLysAlaThrTyrSer 336
Db 1865 GCGCTCCAGCACTGCGCGCACACCGACGAGCGCGTCCACGAGCTGACCTACCGC 1924
QY 337 PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1925 AACCTCGGCACCCAGGAGCTACGCTGAGCTGACGTGACGCGCGCGCGCGCGCGCGCGCG 1984
QY 354 -----AlaProAlaSer 357
Db 1985 AAGCGCGCGCGCGCGCGCTTCTACGCTGGCGCGCACACCGGTGACCGTCCCGCGCGCG 2044
QY 358 ThrThrAlaSerValThrLeuValAlaAsnAspLeu----- 370
Db 2045 GGCAGCGCTCGTGCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2104
QY 371 -----ValIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2105 TACTCGCGGTACGTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2146

RESULT 14

US-09-000-016-1
; Sequence 1, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:

APPLICANT: Akira ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
STREET: 2033 K Street, Lind & Ponack, L.L.P.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores:
Pred. No.: 1,43e-25 Length: 2809
Score: 346.00 Matches: 129
Percent Similarity: 42.00% Conservatives: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.40% Indels: 84
DB: 3 Gaps: 13
US-09-985-689A-1 (1-434) x US-09-000-016-1 (1-2809)
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACAGCTCGCTCGCGGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1009
QY 22 GlnGlyGlnIleValAlaAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGCGTCTGACACCGCGGTGTCGACGAGC----- 1054

QY 42 MethHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCGGACCTGAAGGGCCGGTGACCGGTCACCAAGAACTTACCGCCCGCGCCGCC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCGGGCCGACGAAGGTGGGCGCACGCGCACCGATCGCTCCATCGCGGGCGGCGCGGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGCAAGTACAGGGCGTGCACCCCGCGCGCATCTCAACGCGCAAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGATCCGGT-----TTCGGCGACGACTCCGCGCATCTCCCGCGGATCGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGGCGCGGAGGGCGCGGAGCTCGTCAAGATGAGCTGGCGGATGGACACACCGGAG 1345
QY 137 aTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACCGACCGCTGGAGGCGGG-GTCCGACAGCTCTCCCGCAGAGGCGGCTCTGCTTCGC 1404
QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
Db 1405 CATCGCGCGCGGCAACGAGGCGCGGAG-----TCGATCGGTTCGCGCGGCGCGGCA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 GCGCGCGCTCACGTCGGCGC-----GTCTGA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216
Db 1486 GCACAAAGGACAAGCTCGCGGACTTCCTCCACCGCGCCCGCTCGGCGACGCGGCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGACGTACCGCTCCCGCGGTGGACATCAAGCGCGCTCGGCGGAGGCGCAACGA 1605
QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGGAGGTGCGGTAGGGACCGCGCGCTACATCACCATCTCCGCGACGTGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGGACCGCGGACGTGCGGGCGCGCGCTCTCTGAAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGACTGGACCTCCCGCACTGAAGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGCAAGTACACCCCGTTCGAGCAGGTTTCGGGCGGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSer----- 328
Db 1816 CAAGGCGTCCAGCAGACCGTATCCGACCGCGCTCGGTGAGCTTCGGGCTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGGCACACCGACGAGCGGTCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1936 CCAGGACGTACGCTGAAGCTGACGTTCGACGCGCACCGCCCAAGGCGCAAGCGGCCCC 1995
QY 354 -----AlaProAlaSerThrThrAlaSe 361
Db 1996 GCGGGGCTTCTTACGCTGGCGGCCACCGGTGACCGTCCCGGCGGCGGCGGCGGCTC 2055
QY 361 rValThrLeuValAsnAspLeuAspLeu-----Va 371

Db 2056 CGTCGACATACCGCGCGCACCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGTA 2115
QY 371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CGTGTGCGCCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGTC 2146
RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRIBITOL
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; STRAIN: <Unknown>
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores:
Pred. No.: 1,43e-25 Length: 2809
Score: 346.00 Matches: 129
Percent Similarity: 42.00% Conservative: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.40% Indels: 84
DB: 4 Gaps: 13

```
US-09-985-689A-1 (1-434) x US-09-514-340-1 (1-2809)
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACAGTCGTCGGCGAGATCGCGCCCAAGCGTGGTCCGCC---GGCTACGACGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGCGTGAAGATCGCTCTCGACACCGGTGTCACACGAGC----- 1054
QY 42 MethHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGGAGCTGAAGCGCGGTGACCGCGTCCAGAACTCACCGCGCGCCCGGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCCGCGCAAGGTGGCGCCACGCCACCTCGCTCGATCGCGGGCGGCAGCGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCCAAGGCAAGTACAAAGGCGTGCACCGCGCGCGGATCTCTCAACGGCAAGTGC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGAGCACTCCGCT-----TTCGGCGAGCACTCCGGCATCTCTCGCGCATGGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGCGCGCGAGCGCGCGCGAGTCTGCTCAACATGAGCTGGCGCGCATGGACACCGGAG 1345
QY 137 aTyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleIle 157
Db 1346 ACCGACCCGCTGAGCGGGCGG-GTCGACAGCTCTCCGCCGAGAGGCGCTCTGTTCGC 1404
QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaL 177
Db 1405 CATCGCGCGCGCAACAGGCGCGCGAG-----TCGATCGGTTCGCCCGCGCGCGGA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCTCTACCTCGCGCGC-----GTCCA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerArgGlyPro---ThrLysAspGlyArgIle 216
Db 1486 CGACAAGGACAAGTCCCGACTCTCTCCACCGCGCGCGCTCGCGCGAGCGCGCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGAGCTCACCGCTCCCGCGTGGACATCACGGCGCTCTCGCGGAGGGCACGA 1605
QY 236 oAspSerSerPheTrpAlaAsnHisAspSerIlyTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGGAGTCGTTGAGGACCGCGCGCTACATGACCATCTCCGGCAGCTCAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GCGCACCGCGCACCTCGCGCGCGCGCTCTCTGAAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGAGCTGACCTCCCGCGCACTGAAGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGCAAGTACACCCCGTTCGAGCAGGTTTCGGCGCGGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer----- 328
Db 1816 CAAGCGCTCCAGCAGACCGTATCGCGCGCGCGTCTCTGGTGGAGTTCGGCGTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGCGCACCGAGCAGCGCGTCTACCAAGCAGCTACCTACCGCAACCTCGGCAC 1935
```

```
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1936 CCAGGACGTCACGCTGAAGCTGACGTGACCGCCACCGCCCAAGGCAAGCGCGCCCC 1995
QY 354 -----AlaProAlaSerThrThrAlase 361
Db 1996 GCGGGCTTCTTCACGCTGGGGCGCCACACGCTGACCGTCCCGCGCGCGCGCGCCTC 2055
QY 361 rValThrLeuValAsnAspLeuAspLeu-----Va 371
Db 2056 CCGTCGACATGACCGCGCGCACACCGCGCTCGGGCGGACGCTGGACGCGCTACTCGCGCTA 2115
QY 371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CGTGGTCCGACCGGGCGCGGCGAGACGGTGC 2146
```

Search completed: July 28, 2003, 08:42:47
Job time : 131.115 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein : nucleic search, using frame_plus_p2n model

Run on: July 28, 2003, 03:48:55 ; Search time 332.884 Seconds
(without alignments)
2689.657 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNPVGPQTFLAIVN 434

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2_1/USPRO_pool/US09985689/tunat_25072003_143033_892/app_query.fasta_1.1166
-DB=Published_Applications_NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -LISTS=bits -STARP=1 -END=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DCCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09985689@cgn_1_1_409._runat_25072003_143033_892
-NCPU=6 -ICPU=3 -NO_MMWP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEVTIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	452.5	20.1	1977	13	US-10-090-624-11

2	416.5	18.5	1236	13	US-10-090-624-2
3	416.5	18.5	1962	13	US-10-090-624-15
4	363.5	16.2	3624	14	US-10-156-761-5701
5	363.5	16.2	9025608	14	US-10-156-761-1
6	346.5	15.4	3417	14	US-10-156-761-3306
7	346.5	15.4	9025608	14	US-10-156-761-1
8	332.5	14.8	1329	10	US-09-974-300-1934
9	307	13.7	4765	13	US-10-090-624-5
10	283	12.6	3743	11	US-09-927-827-29
11	282	12.6	3788	11	US-09-927-827-33
12	276	12.3	1306	10	US-09-966-921A-5
13	276	12.3	1330	10	US-09-966-921A-5
14	270	12.0	3303	14	US-10-156-761-5384
15	256	11.4	1485	10	US-09-974-300-1938
16	253	11.3	1971	10	US-09-974-300-1935
17	251	11.2	840	14	US-10-209-812-1
18	242.5	10.8	3884	11	US-09-927-827-34
19	240	10.7	1140	8	US-08-322-678-11
20	240	10.7	1140	8	US-08-322-678-12
21	237.5	10.6	1497	8	US-08-322-678-6
22	237.5	10.6	1497	9	US-09-060-854B-1
23	237.5	10.6	1497	14	US-10-033-325-1
24	237.5	10.6	1497	14	US-10-228-572-1
25	235.5	10.5	1494	14	US-10-104-693-1
26	232	10.3	3452	11	US-09-927-827-30
27	231.5	10.3	1332	14	US-10-156-761-5689
28	227.5	10.1	1140	10	US-09-920-118-13
29	226.5	10.1	1457	14	US-10-202-339-1
30	225.5	10.0	4198	9	US-09-891-711-5
31	225.5	10.0	4338	9	US-09-891-711-3
32	223.5	9.9	1074	9	US-09-837-235-14
33	223.5	9.9	2760	14	US-10-328-459-1
34	222	9.9	522	11	US-09-824-893A-17
35	220	9.8	522	11	US-09-824-893A-14
36	218.5	9.7	1146	10	US-09-920-118-15
37	218.5	9.7	1206	14	US-10-156-761-2788
38	218	9.7	522	11	US-09-824-893A-98
39	215	9.5	522	11	US-09-824-893A-112
40	214	9.5	522	11	US-09-824-893A-15
41	213	9.5	522	11	US-09-824-893A-4
42	213	9.5	522	11	US-09-824-893A-7
43	213	9.5	522	11	US-09-824-893A-9
44	213	9.5	522	11	US-09-824-893A-12
45	213	9.5	522	11	US-09-824-893A-13

ALIGNMENTS

RESULT 1
US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERHYPERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090, 624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

PROV DP
NO-THIS DVA
NO-THIS DVA

OTHER INFORMATION: Synthetic									
US-10-090-624-11									
Alignment Scores:									
Pred. No.:	2.42e-40	Length:	1977						
Score:	452.50	Matches:	138						
Percent Similarity:	44.66%	Conservative:	67						
Best Local Similarity:	30.07%	Mismatches:	153						
Query Match:	20.14%	Indels:	101						
DB:	13	Gaps:	18						
US-09-985-689A-1 (1-434) x US-10-090-624-11 (1-1977)									
QY	8	ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla	27						
DB	433	ATAGGGGCGGATACCGTCTGGAACCTCCCTCGCTACGACGAGCGGTGTGGTGGTGC	492						
QY	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47						
DB	493	ATCGTCGATACGGGTATAGACCGGAAC-----CACCCCGATCTGAAG	534						
QY	48	GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn	66						
DB	535	GGCAAGGTCATAGGCTGTGACGCGCTCAACGCGAGTCCGACCCCTACGATGACCCAG	594						
QY	67	GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys-----	84						
DB	595	GGACACGGAACCCAGTTCGGGTATCGTTCGCGGAACCGGACGCTTAACCTCCAGTAC	654						
QY	85	---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly	101						
DB	655	ATAGCGCTCGCCCGCGCGAAGCTCGTCGGGCTCAAGGTCTCTCGGTGCGCGGTTCG	714						
QY	102	GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla	121						
DB	715	GGAAAGCTCTCACCATCATCGCGGTGTGTGACTGGGTCTCCAGAACAGGACAAAGTAC	774						
QY	122	GlyAlaArgIle-----HisThrAsnSer	129						
DB	775	GGATAAGGTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACCGACTCC	834						
QY	130	TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspTyrVal	149						
DB	835	CTCAGTCAGCGCGTCAACACGCTCGGAGCC-----	867						
QY	150	ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr	169						
DB	868	-----GGTATAGTAGTCTCGTTCGCGCGCGCAACAGCGCGCGCAACCTACACC	918						
QY	170	IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg	189						
DB	919	GTCGGTCAACCCCGCGCGGAGCAAGGTCAATACCGTCTCGTGCA-----	963						
QY	190	ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly	209						
DB	964	-----GTTGACAGCAACGACATCAATCGCCAGCTTCTCCAGCAGGGA	1005						
QY	210	ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer	229						
DB	1006	CCGACCGCGGAGGAGCTCAAGCCGGAAGTCTCGCCCGCGGCTTGACATCATAGCC	1065						
QY	230	AlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAla	249						
DB	1066	CCGCGCGCGGAGC-----GGAACCAAGCATGGGACCCCGGATAAAGACTACTACACC	1116						
QY	250	TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu	268						
DB	1117	AAGGCCCTCTGGAACCAAGCATGGCCACCCCGGACAGTTCGGGCGGTTCGCGCTCATCTC	1176						
QY	269	ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys	286						
DB	1177	CAGGCCAC-----CCGAGCTGGACCCCGGACAAAGGTGAAG	1212						

QY	287	AlaAlaLeuLeaGlyAla-----AlaAspIleGlyLeu	298
DB	1213	ACCGCCTCATCGAGACCGCCACATAGTCCGCCCAAGGAGATAGCGGATCCGCTAC	1272
QY	299	GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu	316
DB	1273	GTTGCG-----GGTAGGTGACGCTCTACAGGCCCATCAAGTAC	1311
QY	317	---AsnValAlaTyrValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyr	335
DB	1312	GACGACTACGCCAAGCTCACCTCACCGGCTCCGTGCGCGACACAGGGAAGCGCCACCCAC	1371
QY	336	SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro	355
DB	1372	ACCTTCGACGTCAGCGCGGCCACCTCGTGACCGCCCTCTACTGGAC-----	1422
QY	356	AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAspLeuValIleThrAlaPro	375
DB	1423	-----ACGGGCTCGAGCGACATCGACCTCTACCTCTACGACCCG	1461
QY	376	AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGly	395
DB	1462	AACGGGACGAG-----GTTGACTACTCTACACCGCTACTAC-----	1500
QY	396	ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGlu	415
DB	1501	-----GCTTCGAGAAGTTCGCTACTACAACCCGCGCGGAACCTGGACGGTCAAG	1554
QY	416	ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn	434
DB	1555	GTCGTCAGCTACAAG-----GGCGGCGCAACTACCAGGTTCGACGTGCGTCAGC	1602

RESULT 2

US-10-090-624-2

Sequence 2, Application US/10090624

Publication No. US20020132335A1

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio

APPLICANT: SHIMOJO, Tomoko

APPLICANT: ASADA, Kiyozo

APPLICANT: KATO, Ikunoshin

TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

FILE REFERENCE: TAKAKURA=6

CURRENT APPLICATION NUMBER: US/10/090,624

CURRENT FILING DATE: 2002-03-06

PRIOR APPLICATION NUMBER: 09/445,472

PRIOR FILING DATE: 1998-12-06

PRIOR APPLICATION NUMBER: 151969/1997

PRIOR FILING DATE: 1997-06-10

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn version 3.0

SEQ ID NO 2

LENGTH: 1236

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Synthetic

US-10-090-624-2

Alignment Scores:

Pred. No.: 1.25e-36

Score: 416.50

Percent Similarity: 42.55%

Best Local Similarity: 30.00%

Query Match: 18.54%

DB: 13

Matches: 141

Conservative: 59

Mismatches: 148

Indels: 122

Gaps: 20

US-09-985-689A-1 (1-434) x US-10-090-624-2 (1-1236)

| QY | 12 | ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly | 24 |
| DB | 24 | GTCGCGCTCAAGTATGCAACTTACGTTTGAACCTGGATGATGATGTTCTTGGAA | 83 |

RESULT 2

US-10-090-624-2
Sequence 2, Application US/10090624
Publication No. US20020132335A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyoko
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
PRIOR FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 1236
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-090-624-2
Alignment Scores:
Pred. No.: 1,25e-36 Length: 1236
Score: 416.50 Matches: 141
Percent Similarity: 42.55% Conservative: 59
Best Local Similarity: 30.00% Mismatches: 148
Query Match: 18.54% Indels: 122
DB: 13 Gaps: 20
US-09-985-689A-1 (1-434) x US-10-090-624-2 (1-1236)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGlu 24
DB 24 GTCGTGACGCTCAAGTTATGCGCAACTTACGTTTGGAACTGGGATATGATGTTCTGGAAT 83

Db 6919455 GTCGACATCCGCGCGCCGCTCCAGCTCGCCCGCGGACCGGCTAC----- 6919502
 QY 245 AspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsn 264
 Db 6919503 -----TACACCTCCATGAGCGGTACGTCGATGCGGACGCCCATGTGCGGGGGTC 6919553
 QY 265 ValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrProLysProSerLeu 284
 Db 6919554 GCGCGCTCCTCCGCGGACGACCGCCGCTGACCGCGCGCGGCTCAAGACCGCGTG 6919613
 QY 285 LeuLysAlaAla-----LeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsn 302
 Db 6919614 ATGTCCACGTCGCGACCACTCGAGCGCTCCGTATATCATGTTGGGGCGGTGCGGTCACT 6919673
 QY 303 GlyAsnGlnGlyTrpGly-----ArgValThrLeuAspLysSerLeuAsnValAlaTyr--- 320
 Db 6919674 GTCCGCGGCGCGCTCGCGCGCGCGCTCACCGCGGACGCGCGGCTCGGCTTCCAC 6919733
 QY 321 -----ValAsnGlnSerSerSerLeuSerThrSerGln 331
 Db 6919734 CGTGGCGCCCATCAGCGCGATCGACCGCTCACGAGAGCGGTCACTACTCCAACTCCCTCC 6919793
 QY 332 LysAlaThrTyrSerPheThr-----AlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
 Db 6919794 GACACGCGGTGCGAGTTGAGCGCTCGCGGTGCGGGCGCGCGCGGTGTCGCCACCCCTC 6919853
 QY 351 TrpSerAspAla-----ProAlaSerThrThrAlaSerValThrLeuValAsnAsp 367
 Db 6919854 GCGACACCGCACTCACTACCGTGGCGCGCCACGCGCGCGCGCGCGCGCGCGCGCGAC 6919913
 RESULT 6
 US-10-156-761-3306
 ; Sequence 3306, Application US/10156761
 ; Publication No. US200301190181
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIRAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156.761
 ; PRIORITY FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 3306
 ; LENGTH: 3417
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(3417)
 US-10-156-761-3306

Alignment Scores:
 Pred. No.: 3,78e-28 Length: 3417
 Score: 346.50 Matches: 137
 Percent Similarity: 43.81% Conservative: 54
 Best Local Similarity: 31.42% Mismatches: 177
 Query Match: 15.42% Indels: 70
 DB: 14 Gaps: 15

US-09-985-689A-1 (1-434) x US-10-156-761-3306 (1-3417)

QY 18 GlyLeuTyrGlyGlnGlyIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
 Db 745 GGGTACGACGCGCAAGGCGTCAAGATCGCGCTTGGACACCGGTGCGAC----- 795

QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
 Db 796 -----CGACCCACCGCGACTCAAGCACCGAGTGGCGGAGTCCCAAGAACTTCTCC 846
 QY 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 847 GCGCGCGCGCGCGCGCGCGCGCTTCGGTTCAGCGCACGCGCGCTCCATCGCGCGCG 906
 QY 78 GlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
 Db 907 GGCACCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 966
 QY 94 PheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThr 113
 Db 967 AACGGCAAGTCTCGACGACACCGGC-----TCGGCGCGCGCGCGCGCGCGCGCTC 1020
 QY 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAl 133
 Db 1021 GGCATGGAGTGGCGCGCGCGCGCGCGCGCTCGTCAACTCGAGCTTGGCGCGCGCG 1080
 QY 133 aValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAsnAs 153
 Db 1081 GACACCGCGCGCGCGCGCGCGCGCTGGAAGGAG-GTCAACAAGCTCTCCGAGGAGAGG 1139
 QY 153 pMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly---GlyThrIleSerAl 172
 Db 1140 CATCTCTTCGCGGTGCGCGCGCGCGCGCGCGCGTTCGGCGCGCGCGCGCGCGCTC 1199
 QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
 Db 1200 CCGCGCGCGCGCGCGCGCGCGCTCACGCTCGCGCGC----- 1236
 QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
 Db 1237 -----GTGAACGACGCGACAGCTGGCTGCTCTCCAGCGCGCGCGCGCGCT 1286
 QY 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
 Db 1287 GGCAGCGCGCGCGCGCGCGCGCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1341
 QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLys-----TyrAl 249
 Db 1342 ----GCGCGCGCGCGCGCGCGCGCTACGACGAGGAGTGGCGCGCGCGCGCGCGCT 1397
 QY 249 aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr 269
 Db 1398 CACCATCTCCGGTACGTGATGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 1457
 QY 269 gGluHisPheValLysAsnArgGlyIleThrProLysProSerLeu-----LeuLysAl 287
 Db 1458 GCAGCAGCAC-----CCCAACTGGTCTGCTTCCCGCGCGCGCGCGCGCGCGCTCA 1496
 QY 287 aAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr 307
 Db 1497 CGCGCTGACCGCGCTCCGCGAAG-----GGCGGCAAGTACACGCGCGTCCAGCAGGCGTC 1550
 QY 307 pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlnSerSerLe 327
 Db 1551 GGGCGGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1610
 QY 327 uSerThrSerGlnLysAlaThrTyrSerPhe-ThrAlaThrAlaGlyLysProLeuLysI 347
 Db 1611 G-----AGCTTCGCGCATCCAGCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1646
 QY 347 iSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA 367
 Db 1647 CAAGCGCGGTCAACCGCAGCTACCTACCGCAACCTCGCGCGCGCGCGCGCGCGCGCTG 1705
 QY 367 sLeuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPhe---- 385
 Db 1706 ACCTCGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCT 1765

QY 386 -----ThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValG 400
 Db 1766 TCGGCGCAGCAAGGTCACGGTCCCGCGGGC-----GCCAAGCCTCGGTGC 1813
 QY 400 lu-----AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValG 417
 Db 1814 ACTTCAGGTCAACAGCAAGCTGGCGGCACCCAGCAGCGCGGTACTCGCGGTACGTGA 1873
 QY 417 InAlaTyrAsnValProValGlyProGlnThrPheSerLeuAla 431
 Db 1874 CGGCCACGGC-----GGCGCCACACCGTGCACGGCG 1908

RESULT 7

US-10-156-761-1/c

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; PRIOR FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 1

; LENGTH: 9025608

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (4187715)

; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Alignment Scores:

Pred. No.: 5,11e-23 Length: 9025608
 Score: 346.50 Matches: 137
 Percent Similarity: 43.81% Conservative: 54
 Best Local Similarity: 31.42% Mismatches: 177
 Query Match: 15.42% Indels: 70
 DB: 14 Gaps: 15

US-09-985-689a-1 (1-434) x US-10-156-761-1 (1-9025608)

QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyArg 37
 Db 4132505 GGTACGACAGCGGCAAGGCGCTCAAGATCGCGTCTCGACACCGGTGTCGAC----- 4132455
 QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
 Db 4132454 -----CGGACCCACCGGACCTCAAGGACCGAGTGGCGGAGTCCCAAGAACTTCTCC 4132404
 QY 58 ArgThrAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 4132403 GCGCGCGCGACCGCGGACCACTTCGGTTCAGGCGACGACGTCGCGTCCATCGCGCG 4132344
 QY 78 GlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
 Db 4132343 GGCACGGCGCGCAAGTCCCAACGCAAGTACAAAGGTGTCGCGCGCGGCGACGATCCTC 4132284
 QY 94 PheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThr 113
 Db 4132283 AACGGCAAGTCTCTGACACACCGGCG-----TCCGGGCGACGACTCCGCGCATCTTGCC 4132230

QY 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTyrGlyAlaAl 133
 Db 4132229 GGCATGGAGTGGCGCGCGGAGCAGCGCGCGAGTCTCAACCTGAGCCTGGCGCGGC 4132170
 QY 133 aValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAs 153
 Db 4132169 GACACCCCGGAGATCGACCCGCTGGAGCGGAG-GTCAACACAGCTCTCCGAGGAGAGGG 4132111
 QY 153 pMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly---GlyThrIleSerAl 172
 Db 4132110 CATCTCTTCGCGATCGCGCGGCAACGAGCGGAGTTCGGCGCACACCATCGGCTC 4132051
 QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAlaLeuArgProSerPh 192
 Db 4132050 CCCGGGCGAGCGCGCGCGCTCACGCTCCGCGC----- 4132014
 QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
 Db 4132013 -----CTGAACGACGCGGACAGCTCGGCTCTTCTCCAGCGCGCGCGCT 4131964
 QY 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
 Db 4131963 GGACGGCGCCATCAAGCGCGGACGTCACCGCACCGCGGTGGACATCACCGCGCGC----- 4131909
 QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLys-----TyrAl 249
 Db 4131908 -----GCCGCGCGCGGCGGCTCATCGACGAGGTCGCGCCAGAGCGGCGGCTACCT 4131853
 QY 249 aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr 269
 Db 4131852 CACCATCTCGGTACGTGATGGCGACCGCGCATCTCGGGCGCGCGCGCATCTCTCAA 4131793
 QY 269 gGluHisPheValLysAsnArgIleThrProLysProSerLeu-----LeuLysAl 287
 Db 4131792 GCAGCAGCAC-----CCCAACTGGTCTGTCGCGGAGTCAAGG 4131754
 QY 287 aAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr 307
 Db 4131753 CGCGCTGACCGGCTCCGCGAAG-----GGCGGCAAGTACACGCGGTTCACAGGGGTC 4131700
 QY 307 pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLe 327
 Db 4131699 GGGCCGTATCGCGTGCACAAAGCGATCAAGCAGTCCGTGATCGCAACCGGACACTCGT 4131640
 QY 327 uSerThrSerGlnLysAlaThrTyrSerPhe-ThrAlaThrAlaGlyLysProLeuLysI 347
 Db 4131639 G-----AGCTTCGGCATCCAGCAGTGGCGGCGCACACGACGA 4131604
 QY 347 leSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA 367
 Db 4131603 CAAGCGCGTCACCCAGCAGCTCACCTACCCCAACCTCGGACGACTGACGTC-ACGCTGA 4131545
 QY 367 spleuAspLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPhe---- 385
 Db 4131544 ACTTCGCGTCGACGGCGGACCAACCCCAAGGCGCTCGCGCGCTCTCTTCAAGC 4131485
 QY 386 -----ThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValG 400
 Db 4131484 TCGGCGCGCAGGAAGTCAAGTCCCGCGCGGCG-----GGCAAGCGCTCGGTGC 4131437
 QY 400 lu-----AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValG 417
 Db 4131436 ACTTCAGGTCAACACGAGCTGGCGGCGCACCGGACCGCGCGGTACTCGCGGTACGTGA 4131377
 QY 417 InAlaTyrAsnValProValGlyProGlnThrPheSerLeuAla 431
 Db 4131376 CGGCCACGGC-----GGCGGCGCAGACCGTGGCGACGGCG 4131342

RESULT 8

US-09-974-300-1934

; Sequence 1934, Application US/09974300

; Patent No. US20020146721A1

; GENERAL INFORMATION:


```

Oy 64 AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsn 83
Db 1204 GATGTCAGGGTCACGGAACCTCACGTAGCTGGAACTGTTGCTGTTACGACAGCAACAAT 1263
Oy 83 ----- 83
Db 1264 GATGCTTGGGATGCGCTCAGTATGCTACTCTGTGTGAATGGGAAGTGTCTCAACACCTCAT 1323
Oy 84 -----LysGlyMetAlaProGlnAlaAsn 91
Db 1324 GCTTGGGATTATACGAACCTTACCAGACACACCGTGCAGGGTGTCTCCAGGTGCCCA 1383
Oy 92 LeuValPheGlnSerIleMetAspSerGlyGlyLeuGlyGlyLeuProSerAsnLeu 111
Db 1384 ATAATGGCAATAAGACTTCTTAGGAGT---GATGGACGGGTAGCATGIGGGATATTATA 1440
Oy 112 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 131
Db 1441 GAAGGTATG---ACATACGCACCAACCCATGTCGACAGCTTATAAGCATGAGTCTCGGT 1497
Oy 132 AlaAlaValAsnGlyAlaTyrThr-----ThrAspSerArgAsn-----ValAspAsp 147
Db 1498 GGA-----AATGCTCCACTACTAGATGTACTGATTCAGAAAGCGTGTCTGGATGAG 1551
Oy 148 TyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGly 167
Db 1552 CTTACCGAAAAGTACGGTGTGTTATTCGTAATAGCTGCAGGAAATGAAGTCTCGCAT 1611
Oy 168 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 187
Db 1612 AACATCGTGGAGTCTGGTGTGTCACAAAGGCAATAACTGTTGGAGCTGCTGCA--- 1668
Oy 188 LeuArgProSerPheGlySerTyrAlaAsp----- 197
Db 1669 GTGCCCAATTAACTGTGGAGTTATGTTTCCCAAGCAGCTTGGATATCTTCATTACTATGA 1728
Oy 198 -----AsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 1729 TTCTATTACTTCCCGCCTACACAAAGCTT---AGAATAGCATCTTCTCAAGCAGAGGG 1785
Oy 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
Db 1786 CCGAGATAGATGTTGAATAAAACCCCAATGTAGTGTCTCCAGGTTCAGGAATTTACPCA 1845
Oy 230 AlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAla 249
Db 1846 TCCTCGCGATGTGGATTGGCGGAGCTGACTTC----- 1878
Oy 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 269
Db 1879 ---ATGCTCGAACTTCGATGGCTACTCCACATGTCAGCGGTGTCGTGCACTCTCATA 1935
Oy 270 GluHisPheValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu 289
Db 1936 AGCGGG---GCAAAGCGCAGGGAATATATACTAATCCAGATATAATTAAGAAGTTCTT 1992
Oy 290 IleAlaGlyAlaAlaAspIle-----GlyLeuGlyTyrProAsnGly 303
Db 1993 GAGACGGGTGCACCTCGGCTTGAGGAGATCCATATATACTAGGCAGAAATACACATGAGCTT 2052
Oy 304 AsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu 323
Db 2053 GACCAAGGTCTGCTTGTAACTGTTACCAAGTCTCTGGGAAATCCTTAAGGCTATAAAC 2112
Oy 324 SerSerSerLeuSerThrSerGln-LysAlaThrTyrSerPheThrAlaThrAlaGlyLy 343
Db 2113 GGCACACACTCTCCCAATTTGATCACTGGGCAGACAAAGTCTCTACACCGACT----- 2164
Oy 343 sProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAlaSerValTh 363
Db 2165 ---TTGCGGAGTACTTGGGTGGACGTTATAAGAGGTCTCTACCAAGGAAC----- 2215

```

```

Oy 363 rLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThr--GlnTyrValGly 382
Db 2216 ---CTATACCTGACATTGCGAGTGGCAGCATTAAGTACGTAGG 2256
Oy 383 Asn-----AspPheThrSerProTyrAsnAspAsnTrp----- 393
Db 2257 GACACGGAGTACAGAACTTTTGAGATCTATGCAACTGAGCCATGGATTAAAGCCTTTTGTG 2316
Oy 394 AspGly-----ArgAsnAsnValGluAsnValPhe-----Ile 404
Db 2317 AGTGAAGTGTAAATCTAGAGAAACAATACCGAGTTGTCCTTAGGGTGAATATGATGTA 2376
Oy 405 AsnAlaProGlnSerGlyThrTyr----- 412
Db 2377 GAGGTCTTGGAGCCAGGTCTCTATGTTGGAAGGATAATCATCATGATCATCAACAAGCCCA 2436
Oy 413 ThrIleGluValGlnAlaTyrAsnValProValGlyProGlnThrPheSer 429
Db 2437 GTTATTGAAGACGAGATCTTGAACACAATGTTATTCCCGAGAGTTCACT 2487
RESULT 10
US-09-927-827-29
; Sequence 29, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 29
; LENGTH: 3743
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE: CDS
; NAME/KEY: CDS
; LOCATION: (1001)..(2743)
US-09-927-827-29
Alignment Scores:
Pred. No.: 5,34e-21 Length: 3743
Score: 283.00 Matches: 127
Percent Similarity: 37.53% Conservative: 55
Best Local Similarity: 26.19% Mismatches: 147
Query Match: 12.59% Indels: 156
DB: 11 Gaps: 24
US-09-985-689A-1 (1-434) x US-09-927-827-29 (1-3743)
Oy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeu-----AspThrGly 36
Db 1502 GCGACGGCGACGCGTCTGCGGTGATTGATACCGCATCACCACTATCGCGACCTCAAC 1561
Oy 37 ArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeu 56
Db 1562 GCCAACATCTCTCGCGGGCTTACGACTTCATCAGCGATGCGACCAACCGCGCATGGCAAC 1621
Oy 57 GlyArgThrAsnAsnAlaAsnAsp----- 64
Db 1622 GCGCGTGACAGCAACGCCGCCAGGAGGACTGTGACGCCCAACGAATCGCGGCC 1681
Oy 65 -----ThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGly 78
Db 1682 GGCATTCCCGCGCCAGCTCCAGCTGGCAGCACCCATGTGCGCGGCACGGTTCGCGGCA 1741
Oy 79 AsnGlySerThrAsnLysGlyMetAlaProGlnAla----- 90
Db 1742 GTGACCAACAACACCAACCGCGGTAGCGGCACCGCTACGGGCCCAAGGTGCTVACCGGTG 1801

```


QY	91	-----AsnLeuValPheGlnSerIleMetAsp	99
Db	1802	CGCGTGTCCGGCAAGTCGGTGCTGCTGCATATACGCCGACGCATCGTCTGGGCC	1861
QY	100	SerGlyGlyLeuGlyLeuProSerAsnLeuGln	117
Db	1862	TCCGGGGCACCGTCACGGCATCCCGCCAATGCTAACCCGGCAGGTGATCAACATG	1921
QY	118	AlaThrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla	137
Db	1922	TGCTCGCGCGCGGTAGCTGCTGCACCACCATGCAGAAC---GCCATCAACGGTGGC	1978
QY	138	TyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeu	157
Db	1979	GTGTGCGCGCGCAC---ACGGTGGTG	2002
QY	158	PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys	177
Db	2003	GTCGGCGCGCACGATGCTCAATGTGCCGT---TCGGTCCGCGCCAACTGCGCG	2059
QY	178	AsnAlaIleThrValGlyAlaThrGluAsn---LeuArgProSerPheGlySerTyr	195
Db	2060	AACGTGATTGGGTGGCGCGCCACCATCTCGCGCGCGGAAGGCCAGCTATTCCAACTTC	2119
QY	196	AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg	215
Db	2120	GGCACCGGTATC---	2131
QY	216	IleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer	233
Db	2132	-----GATGTGTCCGCGCGCGCTGCTGCATCTGTCACACTCAACACGCGCAC	2182
QY	234	LeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly	253
Db	2183	ACCAGCGCGGTAGC-----GCCAGCTATGCCCTCTACACGCG	2221
QY	254	ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal	273
Db	2222	ACCTCGATGGCTCCCGCATGTGGCGCGGTGGTCGGCTGGTCAGTCG-----GTC	2275
QY	274	LysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys-----AlaAlaLeu	289
Db	2276	GCCCGCAGCGCTGACGCCACCGCGGTGGAACTTTGTAAGAACACCGCGCGTGT	2335
QY	290	IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArg	309
Db	2336	TTACCGCGCGCTGC-----TCGGCGCGCTGCGGTGCCGCATC	2374
QY	310	ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSer	325
Db	2375	GTCACGCGCATGCCGCGTCACTGGCGC-----ATCAATGGCGGAGCGCGCGTGGC	2431
QY	326	-----SerLeuSerThrSer	330
Db	2432	GGTGTGGAAACACCTTGACCAAGGCACCTCGGTGACCGCGCTGGCGCGGACTGGC	2491
QY	331	GlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuVal	350
Db	2492	GCGGAATTGAATACACCATCACCGTCCCGCGCGC-----	2527
QY	351	TrpSerAspAlaProAlaSerThrThrAlaSerValThrLeu-----ValAsnAsp	367
Db	2528	-----ACGGGCACCTTGACGTGACCAACACCGCGCGCGCGCGAT	2569
QY	368	LeuAspLeuValIle-----ThrAlaProAsnGlyThrGlnTyrValGlyAsnAsp	384
Db	2570	GCGGACCTGTATGTGGCGCGCGCATGTGCACCGACCGACTCGCGCTTACACC-----	2620
QY	385	PheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIle	404
Db	2621	-----TGCGCGCCATACCGCAC-----GGCAATGCCGAGACTGTCACCATC	2662
QY	405	AsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGly	424


```
Db 2201 GTCAATCTAGCTGGCGGCTACTTCGATCCGAAAGCTACGGCTGTGGCTTC----- 2254
QY 137 AlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle 156
Db 2255 -----ACGGCGTTATCAACAGAGTTCGGCGGTTATGCGCGCAGGGGTACTCGTA 2305
QY 157 LeuPheAlaAlaGlyAsnGluGly-----ProAsnGlyGlyThr--- 169
Db 2306 GTGGTGGCGCGCGCAACGAAGGCTTGGCTGGCTGATGCAGACGACGGCGGCACCTAT 2365
QY 170 -----IleSerAlaProGlyThrAlaLysAsnAlaIleThrVal 182
Db 2366 CCGCGCAACATGATCTGTCAGTCAGCGATCCGGCAATCTGGAGGACGCGATCTCGTGTG 2425
QY 183 GlyAlaThrGluAsnLeuArgPro---SerPheGlySerTyrAlaAspAsnIleAsnHis 201
Db 2426 GGATCGGTGCACAAAGACGACCGCGCAATACGGC----- 2461
QY 202 ValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMet 221
Db 2462 GTGTCGTATTTTCATCCCGCGCGCGACCGCGATGGCGCTCCAAACCTGATGTGTC 2521
QY 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 241
Db 2522 GCCCGCGCGAAAGATCCTGCTCCGCTTACTACGGCTTCGACCGC----- 2566
QY 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 261
Db 2567 CGCGACCCCTCCAGCCTGATGTCGAGATCAGCGCGCACCATGATGCGCGCACCGCATGTG 2626
QY 262 AlaGly-----Asn 264
Db 2627 TCCGGCGTCTGCCCGGGTTTTATCCGCACGCCGCGAGTTCATCGGCTTCCGACCGG 2686
QY 265 ValAlaGlnLeu-----ArgGluHisPheValLysAsn 275
Db 2687 GTCAAGCAACTGCTGCTGCACACCTGCACGACCTGCAGCGCGATCGTTACGTGCGAGGGC 2746
QY 276 ArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAsp 295
Db 2747 AGGGGGGTG-----CCGAATTGTATCGCG-----ATGCTTGGAGAGACGTGA 2788
QY 296 IleGly-----LeuGlyTyrProAsnGlyAsnGlnGly 306
Db 2789 TTCGGAGTGGGGATTGGGATTCGCAACACGCGGTGGTGGC 2830
```

RESULT 12

```
; Sequence 1, Application US/09966921A
; Patent No. US20020103100A1
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Lassen, Soren
; APPLICANT: Pedersen, Poul
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides Having Proteolytic Activity
; FILE REFERENCE: 10097.200-US
; CURRENT APPLICATION NUMBER: US/09/966,921A
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1306
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (50)..(1303)
; OTHER INFORMATION:
; NAME/KEY: mat_peptide
; LOCATION: (371)..()
; OTHER INFORMATION:
; US-09-966-921A-1
```

```
Alignment Scores:
Pred. No.: 6,66e-21 Length: 1306
Score: 276.00 Matches: 100
Percent Similarity: 44.59% Conservative: 40
Best Local Similarity: 31.85% Mismatches: 110
Query Match: 12,28% Indels: 64
DB: 10 Gaps: 17

US-09-985-689A-1 (1-434) x US-09-966-921A-1 (1-1306)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 446 GCGACGGAAATTAAGTAGTACTGTTTAGATACAGGGGTTTATACAGC----- 493
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 494 -----CATTTAGATTAGCTGCTCTCGCGAGCAATGCAAGGATTTTACCAATCTAAT 547
QY 61 -----AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySer 75
Db 548 CCTTTAGTAGATGGTTTCATGCCCGCATGCCAAGGCGATGTTACACATGTTGCGGAAC 607
QY 76 ValLeuGlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsn 91
Db 608 GTATTGGCGCATGCGAGCGCAATAATGCACAAGGCGTTACGGGTGGCTCCGCAAGCGAAA 667
QY 92 Leu---ValPheGlnSerIleMetAspSerGlyGlyLeuGly---GlyLeuProSer 109
Db 668 CTATGGCATATAAAGTATTAGGAGATAACGCGGAGTACTCTGATGATATTGCGAGCA 727
QY 110 AsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsn--- 128
Db 728 GCTATCACACATGTAGCTGATGAGCTTCACGTACAGGTTCCAAAGTAGTAATTAATATG 787
QY 129 SerTrpGlyAlaAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyr 148
Db 788 TCGTAGTGTTCATCTGCCAAGGATTCATTGATTGCT-----AGTCAGTAGATTAT 838
QY 149 ValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGly 168
Db 839 GCATATGGAAGAGGTGTTAATTCGTCGCGCTGGTAAATAGTGGTCGAGCAGCAAT 898
QY 169 ThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeu 188
Db 899 ACAATCGGCTTTCCTGGCGGGCTTCTAATGAGTCGCGCATCGCGCATGGAGAAATGTT 958
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerArg 208
Db 959 CAGCAAAAT---GGAACTTAT-----CGAGTAGCTGATTCTCATCTAGA 1000
QY 209 Gly---ProThrLysAspGly-----ArgIleLysProAspValMetAla 222
Db 1001 GGGATCCGCGCACTGCTGGAGATTATATCATTCAGAGCGGTGATATTCAAGTTTCAGCT 1060
QY 223 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrpAla 242
Db 1061 CCGGAGCAAGTGTA-----CAGTCTACATGG--- 1087
QY 243 AsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 262
Db 1088 ---TACACTGGCGGTTAATATACGATCAGCGGTACATCAATGCTGCTCATGTAGCT 1144
QY 263 GlyAsnValAla-----GlnLeuArgGlu 270
Db 1145 GGGTAGCTGCTAAATCTGTCACCGAATACTTCAATTAAGTCATAGCACTGCGCACA 1204
QY 271 HisPheValLysAsnArgGlyIleThrProLysProSerLeuLysAlaAlaLeuIle 290
Db 1205 GAA---TTGCAAAATCGCGCT-----AAAGTATATGATATTAAAGGTGGTATCGGA 1252
QY 291 AlaGlyAlaAlaAsp-----IleGlyLeuGlyTyrPro 301
Db 1111 111 1111111111
```

```
Db 1253 GCGGAACAGCTGACGATTATGATCATCAGGGTTCGGATATCCA 1294
RESULT 13
US-09-966-921A-5
; Sequence 5, Application US/09966921A
; Patent No. US20020103100A1
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Lassen, Soren
; APPLICANT: Pedersen, Poul
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides Having Proteolytic Activity
; FILE REFERENCE: 10097 200-US
; CURRENT APPLICATION NUMBER: US/09/966,921A
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1330
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-966-921A-5
Alignment Scores:
Pred. No.: 6,85e-21 Length: 1330
Score: 276.00 Matches: 100
Percent Similarity: 44.59% Conservative: 40
Best Local Similarity: 31.85% Mismatches: 110
Query Match: 12.28% Indels: 64
DB: 17 Gaps: 17
US-09-985-689A-1 (1-434) x US-09-966-921A-5 (1-1330)
QY 21 GlyClnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 406 GGCAGCGGAATTAAGGTAGCTGTGTTAGATACAGGGGTTATACAAAGC-----453
QY 41 SerMethHisGluAlaPheArgGlyLysIleThrAlaLeuAlaLeuGlyArgThrAsn 60
Db 454 -----CATTTAGATTAGCTGGTCTGCCAGCAATGCAAGGATTTTACCAATTAAT 507
QY 61 -----AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySer 75
Db 508 CTTTGTAGATGCTTCATGCACCGATCGCCAAAGGCGATGTACACATGTTCGCGAACT 567
QY 76 ValLeuGlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsn 91
Db 568 GTATTGGCGCATGGAGCGAGTAATGGACAAGCGGTTTACGGGGTGGCTCCGCAAGCAAA 627
QY 92 Leu---ValPheGlnSerIleMetAspSerGlyGlyGlyLeuGly---GlyLeuProSer 109
Db 628 CTATGGGCATATAAGATTAAGGAGATAACGCGCGGATCTCTGATGATATTCGACGA 687
QY 110 AsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsn--- 128
Db 688 GCTATCAGACATGTAGCTGTAGTCAAGCTTCACGTACAGGTTCACAAAGTAGTAATAATG 747
QY 129 SerTrpGlyAlaAlaValAsnGlyAlaTyrThrAspSerArgAsnValAspAspTyr 148
Db 748 TCGTAGTCTTCATCTGCCAAGGATTCATTGTTGCT-----AGTCAGTAGATAT 798
QY 149 ValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGly 168
Db 799 GCATATGAAAGGTTGTTAATCTGTCGGCTGGTGAATAGTGGTCAGCAGCAAT 858
QY 169 ThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValClyAlaThrGluAsnLeu 188
Db 859 ACAATCGCTTTCCTGGCGGGCTTGAATGAGTCAGTGGCAGTACGGCATTTGGAGAATGTT 918
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleHisValAlaGlnPheSerSerArg 208
Db 919 CAGCAAAAT---GGAACITAT-----CGAGTAGCTGATTTCATCTAGA 960
QY 209 Gly---ProThrLysAspGly-----ArgIleLysProAspValMetAla 222
Db 961 GGGAAATCCGGCAACTGCTGGAGATTATATCATTAAGACGGTGATATTGAAGTTTCACCT 1020
QY 223 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 242
Db 1021 CCGGAGCAAGTGTA-----GAGTCTACATG--- 1047
QY 243 AsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAla 262
Db 1048 ---TACACTGCGGTTATAACGATCAGCATCAGCGGTACATCACTGCTACCTCATGTAGCT 1104
QY 263 GlyAsnValAla-----GlnLeuArgGlu 270
Db 1105 GGGTTAGCTGCTAAATCTGGTCAGCAATACTTCAATTAAGTTCATACCACTGCGGACA 1164
QY 271 HisPheValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaLeuIle 290
Db 1165 GAA---TTGCAAAATCGCGCT-----AAAGTATATCATATTAAGGTGATATCGGA 1212
QY 291 AlaGlyAlaAlaAsp-----IleGlyLeuGlyTyrPro 301
Db 1213 GCGGAACAGGTGACGATTATGATCATCAGGGTTCGGATATCCA 1254
RESULT 14
US-10-156-761-5384
; Sequence 5384, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5384
; LENGTH: 3303
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3303)
US-10-156-761-5384
Alignment Scores:
Pred. No.: 1.25e-19 Length: 3303
Score: 270.00 Matches: 112
Percent Similarity: 37.61% Conservative: 58
Best Local Similarity: 24.78% Mismatches: 168
Query Match: 12.02% Indels: 114
DB: 17 Gaps: 17
US-09-985-689A-1 (1-434) x US-10-156-761-5384 (1-3303)
QY 68 HisGlyThrHisValAlaGlySerValLeuGlyAsn-----GlySerThrAsnLys 84
Db 1168 CACGCGACGCGCTCGCGGCATCAGCGCGCAAGCCCTGTTTCGGCGGCAAGATGAAC 1227
QY 85 GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGly--- 103
Db 1228 GCGCGCGCGCGCGCGGCGAAGATGCTCTCTCGCGTCCCTGCACTGCACTGCGCGGCGGCTGC 1287
```

QY 103 ----- 103
Db 1288 ACCAAGCTGCGCTCACCAGGCGCATGACGACCTCGTCCCAACCGTGGCGTGACATC 1347
QY 104 -----LeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
Db 1348 GTCAACATGTCATCGCGCGTCTCGCGG-----CTGAACAGCGGCAAC 1392
QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 139
Db 1393 AACGCGCGCGAGCTCTACAG----- 1416
QY 140 ThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeuPheAla 159
Db 1417 -----CGTCTCATGACACCTAC-----GGCGTCCAGCTGGTGATCTCC 1455
QY 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
Db 1456 GCGGCAACTCGCGCCCGCGGCCAACACCATCGCGCGCCGCTGCCGCAACAAGTC 1515
QY 180 IleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGlySerTyrAla 196
Db 1516 ATCTCGTGGCGCGGCCATCTCCAAAGCAGACCTGGCGCGCCCACTACGGCTCCCGAGGTG 1575
QY 197 AspAsnIleAsnHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIle 216
Db 1576 GAGAAGCGGTACGCGCATGCGGTCTCTCTCGCGCGCGCGGTGAGGACGCGGCTTC 1635
QY 217 LysProaspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeu----- 234
Db 1636 ACGGCGAGCTGACCGCGCCCGCGCCCTCGATCAACTCCACGACACCTGGCTGGCGGGC 1695
QY 235 AlaProaspSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 254
Db 1696 TCCCGGTGCGCGAGCGGCTACTGCTCGCGCGCGCTACTCGATCTCCAGGCGACG 1755
QY 255 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 274
Db 1756 TCGATGGCGTCCCGCGAGCGCGCGCGCTTCGCGCGCTGCTGAGC---GCCGCCAAG 1812
QY 275 AsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 294
Db 1813 CACAAGGCGTACGAGCTACCGCCCGCCACCTCGCGCGCGCGCTCGACCGCGCAG 1872
QY 295 Aspile-----GlyLeuGlyTyrProAsnGlyAsnGlnGly 306
Db 1873 CACATCAAGGTGTGTCAGCGGTACGAGGCGCGCGCGCTCATCAACATCGTGGACGCG 1932
QY 307 TrpGlyArgVal----- 310
Db 1933 TGGAGTCCATCAAGGAGGACGCCACGCGCGCACGACTACACCGTGAAGCGCGCGTGCAC 1992
QY 311 ---ThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSer 324
Db 1993 ACCGCGATCGACGACGCGTTGAAGACGCGCGGCTTCGCGCGCGCGCTGACGCGCGAG 2052
QY 325 SerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr----- 340
Db 2053 GCGCGGCTCAAGCGCGGTGACAAAGACGTACGACGTCAACATCAGCGCGTCCGCGG 2112
QY 341 AlaGlyLysProLeuIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
Db 2113 GCCGACGCGCGATCCGCGACGAGCTGTACTTCGAGAACAAACGCGCGCGTACCTCCGG 2172
QY 361 SerValThrLeuValAsnAspLeuValIleThrAlaPro----- 375
Db 2173 ATCGTCGCGC---TCGACGAGGTCAAGCTCCCGCTGAACGCGCGGTGACCGTCAAGTGC 2229
QY 376 -----AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProtyr 389
Db 2230 CAGGCGCGCGCGAGTGGCGCGCTGAAGAGCGCGATCTCGAGGTCGAGCAGCGCGC 2289
QY 390 AsnAspAsnTrpAspGlyArgAsnValGluAsnValPheIleAsnAlaProGlnSer 409

Db 2290 ACCGAGGCGCGTGCAC---AAGCAGATCTGTGACCGTCTGCTCCGCGCGTCAAG 2346
QY 410 GlyThrTyrThrIle-----GluValGlnAla-----TyrAsn 420
Db 2347 TACAGTACTCGCGTCTCGGTTCGTGAGCGCAACAGCAGCACCAGTCGTACTTCGTGACG 2406
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIle 432
Db 2407 GTCCCCGAGGCGCGCAAGTCCCTCGAGGTCCGCGATC 2442

RESULT 15

US-09-974-300-1938
; Sequence 1938, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1938
; LENGTH: 1485
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1938

Alignment Scores:

Pred. No.: 1.38e-18 Length: 1485
Score: 256.00 Matches: 94
Percent Similarity: 42.82% Conservatives: 52
Best Local Similarity: 27.57% Mismatches: 123
Query Match: 11.39% Indels: 72
DB: 10 Gaps: 16

US-09-985-689a-1 (1-434) x US-09-974-300-1938 (1-1485)

QY 11 AspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlnIleValAlaValAlaAsp 30
Db 573 GATGCTCGGAAAGAACGCGTAT-----AAAGCAAGAGAGCGCGTCTCAATTGAC 626
QY 31 ThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGly----- 48
Db 627 ACCGCGGTAGATGG-----GATCATCCGCGGTGAAACAAAGTACAGAGGTACAC 680
QY 49 -----LysIleThrAlaLeuTyrAlaLeuGlyArgThr 59
Db 681 CCTTCTAAACCGAACAGTCCAGACATGAGTTCAACTGGTTGATCGGTGAAGCAATAAG 740
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 741 AAACGCCATATGATGAGTTGGGCCCGCACACATGTGACAGCAGCATGTCGCGTCA 800
QY 80 -----GlySerThrAsnLysGlyMetAlaProGlnAlaAsn---LeuValPheGln 95
Db 801 GAAGCGCGCGCAAAACACAGATCGGTGTCCGCCGACGCGGCAAAATGGATGAGTGTCA 860
QY 96 SerIleMetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPhe 115
Db 861 GCCTTTTCAGAGACGCGGAGATGAAAAATCCCTTTTGGCTGCAGGAGATGATTTTG 920
QY 116 SerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
Db 921 GCCCGCAAGATGCGAAAGGA---AAAGCGCACCTCTGAAAGGCTCCGATGTCGTCAAC 977

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:57:50 ; Search time 16.0185 Seconds

(without alignments)

2605.570 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAOSSVGLY.....EVQAYNVPVGPQTFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	523.5	23.3	1743	2 T18279	multidrug resistan
2	497	22.1	1905	2 T18267	multidrug resistan
3	347.5	15.5	444	2 B83691	intracellular alka
4	320.5	14.3	442	2 A69587	intracellular alka
5	310.5	13.8	806	2 A41341	microbial serine p
6	302.5	13.5	1398	2 T28159	pyrolysins (EC 3.4.
7	283	12.6	580	2 S11890	serine proteinase
8	281	12.5	419	1 S25835	subtilisin (EC 3.4
9	280	12.5	799	2 G83753	subtilisin-type pr
10	279.5	12.4	1345	2 T29090	surface layer-asso
11	277	12.3	420	1 S23407	subtilisin (EC 3.4
12	267	11.9	715	2 JC4908	alkaline serine pr
13	263	11.7	1331	2 A72647	probable surface l
14	259.5	11.5	757	2 C84120	subtilisin-type pr
15	257.5	11.5	513	1 A35742	aqualysin (EC 3.4.
16	253.5	11.3	894	2 F69730	cell wall-associat
17	251	11.2	627	2 D75392	serine proteinase,
18	246.5	11.0	402	1 JU0332	alkaline proteinas
19	246.5	11.0	534	1 JS0173	alkaline proteinas
20	243	10.8	519	2 S71451	halolysin R4 (EC 3
21	242.5	10.8	401	2 I39974	serine proteinase
22	240	10.7	380	2 A49778	high-alkaline seri
23	240	10.7	488	2 A11930	proteinase (import
24	235.5	10.5	382	1 SUBSN	subtilisin (EC 3.4
25	235	10.5	378	2 A33973	high-alkaline seri
26	235	10.5	382	2 I39780	subtilisin (EC 3.4
27	234	10.4	910	2 C69456	subtilisin sendai
28	234	10.4	1374	2 D72593	hypothetical prote
29	232.5	10.3	525	2 G84406	halolysin (importe

30 231 10.3 321 1 S27501 alkaline proteinas
31 230.5 10.3 379 1 SUBSCL subtilisin (EC 3.4
32 229.5 10.2 601 2 JC4576 serine proteinase
33 227 10.1 1167 1 A35066 streptococcal C5a
34 226.5 10.1 1118 2 H97298 subtilisin like pr
35 225.5 10.0 381 2 JH0778 subtilisin (EC 3.4
36 225.5 10.0 613 2 S75976 hypothetical prote
37 225.5 10.0 1052 2 T17093 intraluminal subti
38 223.5 9.9 381 1 SUBSS subtilisin (EC 3.4
39 223.5 9.9 381 1 SUBSI subtilisin (EC 3.4
40 223.5 9.9 381 2 J01487 subtilisin (EC 3.4
41 222.5 9.9 275 2 JC1085 subtilisin (EC 3.4
42 222 9.9 384 2 JC4802 alkaline proteinas
43 221.5 9.9 272 2 A23624 subtilisin (EC 3.4
44 221.5 9.9 530 2 A42605 halolysin (EC 3.4
45 221 9.8 1036 2 JC5568 serine proteinase

ALIGNMENTS

RESULT 1

T18279

multidrug resistance transport protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18279

R:Shaulsky, G.; Loomis, W.F.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z18855

A:Accession: T18279

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1743 <SHA>

A:Cross-references: EMBL:060086; NID:g1399914; PID:g1399915; PIDN:AAB03331.1
C:Genetics:
A:Gene: tagC

Query Match

Best Local Similarity 23.3%; Score 523.5; DB 2; Length 1743;

Matches 165; Conservative 81; Mismatches 150; Indels 195; Gaps 24;

QY	19	LYGGOIVAVADTGLDTR	---NDS-----SMHEAFRGKITALYALGRINNANDNGH	68
Db	314	LRGKGQLSIADTGLDGS	HCFFSDSKYPIPLNSVNLNHR-KVVTYITSTSDSDKVDGH	372
QY	69	GTHVAGSVLG	-----NGSTKGMAPQANLVFQSIMDSGGGLGL--PSNLOTLFQAY	119
Db	373	GTHICGSAAGTPE	DSVNTSFSGLATDAKIAF---FDLASGSSSLTPPSDLKQLOPLY	429
QY	120	SAGARIHTNSGA	---AVNGAYTTDSRNVDYVRKN-DMTILFAAGNEGPNGTIS--A	172
Db	430	DAGARVHCDWSG	SVSEGYTGSYSSDTASIDDFLTHPDPIILRAAGN---NEQYLSLT	486
QY	173	PGTAKNAITVGATENLR	-----PSFGSYADNI-----	199
Db	487	QSTAKNVTVGARHQT	IHENYLTGDPNYINVQSSVDINQELICDFDSRYCNYTTAQCCLES	546
QY	200	-----	-----NHVAFSSRGPTKDGRIKPDVMAQPFIL	228
Db	547	NATTGLASCCPTLL	RKKSVIDAANTOPLYNENNICSFSKGPPTHDGRMKPALVAPGEIT	606
QY	229	SARSLSA	-----PDSSFWANHDSKYAYMGSTMATPIVAGNVAQREH-----F	272
Db	607	SARSINGANT	TCQCGSL-PTINALLA-ISGISMATSFAAAATTIILRQYLVGDGYPTGSI	664
QY	273	VKNRGITPKP	LLKKAALIAGA-----ADILGYPNGN-----QMGGRVT	311
Db	665	VESKRLQPTG	SLKALMINNAQLLNGTFOITSSITYPSNOVFENFACASLVQHGAIK	724
QY	312	LKSLNVAYN	NESS-----	338
Db	725	MSNWLVVNN	NNNNNTSDGITKFDGIGGLDLRLVKPNOWKEESLSTGQNTSYCFTYK	784

QY 339 -----ATACKPLK---ISLWVSDAPASTTASVTIVNDLVL-----TAPNGT 378
Db 785 PSSSSNSGNIPRVVATLWTDPPSYAGAKFNVLNLDLTMIIYRDNGSTFTFYSNOGGS 844
QY 379 QYVGNDFTPYNDWDRNVNENFINAPQSGTITIEQAVNPVGPQTES 429
Db 845 SFLG---LAPQDT---LNNVEGIVHNPTEPMTYRFVWAGTVNMPGPNFS 889

RESULT 2
T18267
multidrug resistance protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18267
R:Shaulsky, G.; Kuspa, A.; Loomis, W.F.
submitted to the EMBL Data Library, January 1995
A:Description: An MDR transporter/serine protease gene is required for prestalk special
A:Reference number: T18850
A:Accession: T18267
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1905 <SHA>
A:CROSS-references: EMBL:U20432; NID:G664839; PID:G664840; PIDN:AAA62212.1
C:Genetics:
A:Gene: tagB

Query Match 22.1%; Score 497; DB 2; Length 1905;
Best Local Similarity 28.0%; Pred. No. 1.3e-23;
Matches 162; Conservative 72; Mismatches 162; Indels 182; Gaps 22;

QY 19 LYGGQIVAVADTGLDTCR---NDS-----SMHEAFRGKITALYALGRTNNDNTNGH 68
Db 376 LRKGQILSIADTGLDGHCFDSKYPPIFNOVNHNRKVT---YIYHDEYVNGH 432
QY 69 GTHVAGSVLNG-----STNKGMAPOANLQFQIMSDSGGLGGLPSNLQTLFSAQYSA 121
Db 433 GTHVCGSAAGTPEDSSWAISSFSGLATDAKIAFYD-LSSGSGSEPTPPEDYSQMYKPLIDA 491
QY 122 GARTHNSWGA---AVNGAYTDSRNVDDYRK-NDMTILFAAGNEGNGGTISAPGTA 176
Db 492 GARVHGDWSGVSLQYGGYSDAGGIDAFLEYEPEFSILRAAGN-NELFASLLAQATA 550
QY 177 KNATITVGTENLRPSFGS-----YADNI-----PNGNOGQGRVTLDKSLNVA 199
Db 551 KNATITVGAETAHVNVSDALEYDESDANFQPCLEDKCYCNYTTAKCCSEVSNVKG 610
QY 200 -----NHVAQFSRGPTKGRKPKPDVWAPGTFILSARSS- 233
Db 611 QLCCPASIKONASDSFTTQPOFYNNENMGFSKSGPTHGRLKPDIVAPGEYITSRSNG 670
QY 234 -----LAPDSSEFWANHDSKYATMGTSMATPIVAGNVAOLREHF-----VKNRGI 278
Db 671 ENSTDQCGDGL--FNANGLSISGTSMATPLATAATTILRQLVLDGYFTGESVENKL 728
QY 279 TPKPSLLKAAALIAGAADIIGLY-----PNGNOGQGRVTLDKSLNVA 319
Db 729 LPTGSLKALMINNAQLNGTVFWSASSTNPSNAIFEQINGANLIQWALRMN---NWL 785
QY 320 YNESS-----SLTSOKAT-----YST-----ATAGK 343
Db 786 YKSSNPTPPSRWIGTGGGLGNOKATFEWKEDSLSSGLNKSYCFYKPSSSSSSGSGGGGT 845
QY 344 P-LKLSLVWSDAPASTTASVTIVNDL-----VITAPN--GTQYVGNDFTPSYND 391
Db 846 PRIVATLWTDPPSYSGAKFNVLNLDLNLNDDSDSIITIGNSGSLQAGKVAQP--- 902
QY 392 NWDGRNVNENFINAPQSGTITIEQAVNPVGPQTES 429
Db 903 ---DTLNNVEGIINPTKAMNYKFTIAGTNVPIGPQKFS 938

RESULT 3
B83891
intracellular alkaline serine proteinase aprX [imported] - Bacillus halodurans (str
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: B83891
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: B83891
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-444 <STO>
A:CROSS-references: GB:AP001513; GB:BA000004; NID:G10174345; PIDN:BA05649.1; GSPDB
A:Experimental source: strain C-125
C:Genetics:
A:Gene: aprX

Query Match 15.5%; Score 347.5; DB 2; Length 444;
Best Local Similarity 29.9%; Pred. No. 6.4e-15;
Matches 107; Conservative 56; Mismatches 110; Indels 85; Gaps 16;

QY 11 DVAOSSYGLYCGQIVAVADTGLDTCRNDSSSMHEAFRGKITALY-ALGRTNNDNTNGH 69
Db 135 EVIRNGETLTGKDVITIAVIDTGI-----YPHEDEGRKAKAFVDFVNOREEPYDDNGHG 187
QY 70 THVAGSVLNGSTN---KGMAPQANLQFQIMSDSGGLGGLPSNLQTLFSAQYSAQARI 125
Db 188 THCAGDAAGNGASSDQYRGPAPENNVIGVKVLNK-QMGSLSEIMGV----- 235
QY 126 HTNSGAAVNGAYTTDSRV-----DDYVR-----KNDMTILFAAGN 162
Db 236 ---EMCIQYNEEHPDPDTHIISMSLGGQALPYNEOEDPMVRIVEEAWNAGNVCVVAAGN 292
QY 163 EGPNGGTISAPCTAKNAITVGADENLRPSFGSVADNHNHVAQFSRGPTKGRKPKPDYMA 222
Db 293 SGPDQATITASPGVSEKVIIVGALDD-RDITDREDD---VAPFSRGPTIYKPKPDILA 348
QY 223 PGTFILSARSSIAPOSSF-----WANHDSKYATMGTSMATPIVAGNVAOLREHFVKNRG 277
Db 349 PGVNVISLRS---PNSFYDKIQKSRVSGSHYTMMSGTSMATPGVAGVVALMLQH---EPN 402
QY 278 ITPKPSLLKAAALIAGAADIIGLYPNGNOGQGRVTLDKSLNVA---AYNESSLSLSQK 332
Db 403 LTPDE--VKTRLM-----EGDRA-----DRDPNVYAGYISAEGAIPNSE 443

RESULT 4
A69587
intracellular alkaline serine proteinase aprX - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: A69587
R:Kunzt, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.;
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.
C.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.;
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lard
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Y. M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Sc
A:Authors: Schleich, C.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.;
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uch
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yosh
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus sub
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69587
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA


```

C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-32/Domain: signal sequence #status predicted <SIG>
F:168-423/Domain: subtilisin homology <SBT>

Query Match          12.6%; Score 283; DB 2; Length 580;
Best Local Similarity 26.2%; Pred. No. 1.2e-10;
Matches 127; Conservative 55; Mismatches 147; Indels 156; Gaps 24;

Qy      21  GQGIIVAVADTGL-----DTGRNDSSMHEAFPGKITALYALGRTNNAND----- 64
      ||| ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db     168  GSGTWVAVIDTGITSHADLNANILAGDYFISDATTARDNGGRDSSNAADGGWYAANECCA 227

Qy      65  -----TNGHGHVAGSVLNGSTNGKMAFQA-----NLVFSQIMD 99

```

```

228 GIPAASSWHGTHVAGTVAAVTNNTTGVAGTAYCAKVVVPVRVLGRCGGSLSDIADAIVWA 287
    ||| :||| : : | :||| :
100 SGGGLGGLPNSLQ--TLFSQAYSAGARIHTNSWGAANYNCAYTTDSRNVDYVVRKNMTIL 157
    ||| :||| : : | :||| :
288 SGGTVSGIPANANPAEVIINMSLGGGSCSTTMQN-AINGAVSRGT-----TVV 334
    ||| :||| : : | :||| :
158 FAAGNEGPNNGTISAPGTAKNALITVGATEN-LRPSFGSYADNINHVAOFSSRGPTKDGR 215
    ||| :||| : : | :||| :
335 VAAGNDASNVSG-SLPANCANVIAVAATTSAGAKASYNFGTGI----- 377
    ||| :||| : : | :||| :
216 IKPDVMAPGTFILSARSS--LAPDSSFWANHDSKIAYMGGTSMATPIVAGNVQAQLREHFV 273
    ||| :||| : : | :||| :
378 --DVSAPGSSILSTLSNGTTTPGS-----ASYASYNGTSMASP HVAGVALVQS-V 425
    ||| :||| : : | :||| :
274 KNRGITPK--PSSLK--AALIAGAADI GLGPNGNOGHGRVTLDKSLNYAYVNES---- 325
    ||| :||| : : | :||| :
426 APTALT PAAEVTELLKNTARALPGAC-----SGCGAGIVNADAAVTAA-INGSGGGG 477
    ||| :||| : : | :||| :
326 -----SLSTSOKATYSFTATAGKPLKLILVWSDA PASTVASVTL---VND 367
    ||| :||| : : | :||| :
478 GGNLTNTGTPVTGLGAATCAELNYTITVPAG-----SGILT VTTSGSGD 523
    ||| :||| : : | :||| :
368 LDLVI---TAPNSTQYVGNDFTPSYNDWGDRNNVENFIPA QSGTYTIEVQA NYDPVG 424
    ||| :||| : : | :||| :
524 ADLYVRAGSAPTDSAVT---CRPYRS-----GNAETCTITAP-SGTYIVRLKAYS---- 569
    ||| :||| : : | :||| :
425 POTFS 429
    ||| :||| : : | :||| :
Db 570 --TFS 572

RESULT 8
S25835
subtilisin (EC 3.4.21.62) precursor - Bacillus sp. (strain TA41)
C:Species: Bacillus sp.
C:Date: 22-Nov-1993 #sequence_revision 20-Feb-1995 #text_change 18-Jun-1999
C:Accession: S25835
R:Davail, S.; Feller, G.; Narinx, E.; Gerday, C.
Gene 119, 143-144, 1992

```

```
A>Accession: S25835
A>Molecule type: DNA
A>Residues: 1-419 <DAB>
C>Cross-references: EMBL:X63533; NID:g40198; PIDN:CAA45096.1; PID:g40199
C>Superfamily: subtilisin; subtilisin homology
C>Keywords: extracellular protein; hydrolase; serine proteinase
F.1-23/Domain: signal sequence #status predicted <SIG>
F.24-110/Domain: propeptide #status predicted <PRO>
F.111-419/Product: microbial serine proteinase #status predicted <MAT>
F.135-373/Domain: subtilisin homology <SBT>
F.144_184_359/Active site: Asp, His, Ser #status predicted
Query Match 12.5%; Score 281; DB 1; Length 419;
Best Local Similarity 33.0%; Pred No. 1.le-10;
Matches 87; Conservative 38; Mismatches 97; Indels 42; Gaps 13;
QY 21 GQQQIVAVADTGDTGRNDSSMH-EAFRCIKITALYALGRT---NNANDTNGCHTHVAGSV 76
| | |||||::: | : | : |
| | |||||::: | : | : |
```


Db 135 GAGINIAVLDTGNTNHPDLSNNVEQCKD-----FTVGTNFTNSCTDROGHGTHVAGSA 189
QY 77 LGNGSTNK---GMAPOANL-VFQSIMDSGGGLG-GLPSNLTQTLFSQAYSAGARLHTN-SW 130
Db 190 LANGGTGSGVYGAVPADLWAKVIGDDSGGYADDIAEAIIRHAGDQATALNTKVIVNMSL 249
QY 131 GAAVNGAYTTDSRNVDYVRKNDMTILFAAGNPGNGGTISAPGTAKNALITVGATENLRP 190
Db 250 GSSGESSLLIT---NAVDAAYDKGVLLIAAAGNSGPKPGSIGYPGALVNAVAVAALENTIQ 306
QY 191 SFGSYADNLNHVAQSSRSGPTKDG-----RKPDYMACTFTILSRSSSLAPDSSFWANH 244
Db 307 N-GTY-----RVADFSRSHKKTAGDYVIOKGDVEISAPCAAVYST-----W-F 348
QY 245 DSKYAYMGTSMATPIVAGNVAQL 268
Db 349 DGGYATISGTSWASPHAGLAAKI 372

RESULT 9
G83753
subtilisin-type proteinase (EC 3.4.21.-) vpr precursor [similarity] - Bacillus halodurans
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83753
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: G83753
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-799 <STO>
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04550.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: vpr
C:Superfamily: microbial serine proteinase vpr; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-29/Domain: signal sequence #status predicted <STO>

Query Match 12.5%; Score 280; DB 2; Length 799;
Best Local Similarity 24.7%; Pred. No. 3e-10;
Matches 117; Conservative 50; Mismatches 135; Indels 172; Gaps 16;

QY 18 GLYGQGIQVAVDTCGLDTRNDSSMHEAFRKGITATYALGRNNDT-----NGH 68
Db 171 GYTCEGITVALDRCVDYTHPD--LVHAF-GDYKQWDFIDNDDPQETPPGDPGRGIETH 227
QY 69 GTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGPSNLTQTLFSQAYSAGARIHTN 128
Db 228 GTHVAGTVAANGLI-KGVAPDANLILAYRVLGPGG--RGSTAGVIAGIERAVQDGADIMNL 284
QY 129 SWGAANGVAYTTDSRNVDYVRKNDMTILFAAGNPGNGGTISAPGTAKNALITVGATENL 188
Db 285 SLGNTLNDPDTATIAL-DWMAECGVAVTSGNSGPNNTVGSPTSRDALSVCAT--- 340
QY 189 RPSFGSY-----ADNINH----- 201
Db 341 RLPNKYKASVFTSDGIDYPSADIMGFPDSDEELLELDGETYEYAFAGLGKPCDFGCVDE 400
QY 202 ----- 201
Db 401 GKIALIVRGELPFVEKAENAKAAGAVGAIYNNVAGVQTPVPLAIPITMLSNEDGLKMR 460
QY 202 -----VAQFSRSGPT-KDGRIPKPDVMAAGFTILSARSSLAPD 237
Db 461 NELENGONTVFTSTFEDKLVGETVADFSRSGPVNMTWMIKPDVSPAGVAIVSTIPTHOPD 520
QY 238 SSTWANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADI- 296
Db 521 DPY-----GYGSRQGTSMASPHVAGAAALLLEAH-FNWGV-----DHVKAALMNTAENLV 569

QY 297 ---GLGYPNGNOGWRVTLDKSLNVAIVNNESSLSTSQKATY-SPTATAGKPLK 346
Db 570 DENGNRYPHTQAG-----SIRIVDALESETLVTPGSRHSGFTFTKERGQVE 617

RESULT 10
T29050
surface layer-associated STABLE proteinase - Staphylothermus marinus
N:Alternate names: hyperthermostable proteinase
C:Species: Staphylothermus marinus
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T29050
R:Mayr, J.; Lupas, A.; Kellermann, J.; Eckerstorn, C.; Baumeister, W.; Peters, J.
Curr. Biol. 6, 739-749, 1996
A:Title: A hyperthermostable protease of the subtilisin family bound to the surface
A:Reference number: Z20559; MUID:96385442; PMID:8793300
A:Accession: T29050
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1345 <NAR>
A:Cross-references: EMBL:U57968; NID:gl374755; PID:gl374756; PIDN:AAB02323.1
A:Experimental source: strain F1
C:Function:
A:Description: probably serves an exodigestive function related to the organism's
A:Note: stoichiometric S-layer component

Query Match 12.4%; Score 279.5; DB 2; Length 1345;
Best Local Similarity 29.9%; Pred. No. 6.4e-10;
Matches 99; Conservative 45; Mismatches 106; Indels 81; Gaps 14;

QY 46 FRGKITAYALGRTNNDTNGHGHVA-----GSVL-----GNSGTNK--GMAPOANLV 93
Db 445 YQRYLAL-----VSDFHGHGTSVATVIRASGRVLYDLYGDGKLYRIMGVAPGAKI- 495
QY 94 FQSIMDSGGGLGPSNLTQTLFSQAYSAG-----ARHNTSW 130
Db 496 -----AGGDALWLLGNILVL--EAWLAGFNITVEEDGYVYVLSLDPPGPHRADILSNW 546
QY 131 GAAVNGAYTTDSRNVD-----DVKRNDMTILFAAGNPGNGGTISAPGTAK 177
Db 547 GSTYINFWLOOFGIDYRSSFMDRLAIRNTLIGDRVTIVFAAGNPGGYSNGAPGTGL 606
QY 178 NAITVGATE--NLRPSFG---SYADNINHVAQSSRSGPTKDGRIPKPDVMAAGFTILSARS 232
Db 607 LVITAGASTLWDYTRIYGYPEGYAD---EVIPFSRSGTGGQYPPDPINIGAFEWASTR 663
QY 233 SLAPDSSFANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAG 292
Db 664 TI-DGRGYGAQPD-----VFGTSEATPYTSGTLALVFOAYKEVYNTTDPPTAKILKSS 718
QY 293 AADIGLGYPNNGOGRVTLDKSLNVAIVNE 323
Db 719 AKDI--WYPAFSGSGRVDKAAADTVFISE 747

RESULT 11

S23407
subtilisin (EC 3.4.21.62) l precursor - Bacillus sp. (strain TA39)
C:Species: Bacillus sp.
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 18-Jun-1999
C:Accession: S23407
R:Narinx, E.; Davail, S.; Feller, G.; Gerday, C.
Biochim. Biophys. Acta 1131, 111-113, 1992
A:Title: Nucleotide and derived amino acid sequence of the subtilisin from the antar
A:Reference number: S23407; MUID:92236481; PMID:1581352
A:Accession: S23407
A:Molecule type: NAR
A:Residues: 1-420 <NAR>
A:Cross-references: EMBL:X62369; NID:g40200; PIDN:CAA44227.1; PID:g40201
C:Genetics:
A:Gene: subtil
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase

QY . 68 -----HGTHVAG-----SVLGNCSSTNKGMAPQANLVFQSIMDSGGGLCGCLP 108

Db 630 GVAKPDVWVIGSAWAG-----VPVLTGLGNSLAFDFGCTSEATPMTSGSVALVISAQY 685
QY 274 KNRGITPKSLKALIAAGADIGLGPNGQGWGRVTLDKSLNV-----AY 320
Db 686 QAFGAKPSGLVKALKLSTARTGA--DAFTQGGQVDYVAVKAVLEGVPIALSTSVY 743
QY 321 VNESSLSSTQKATSFATATACKPLKISLWSDA--PASITASVTLVNDL 368
Db 744 ENVYSLLS-----GYSYPLAPNPVEDTQIYPGVLKPGETAVETLVKLTL 798

RESULT 14

C84120
subtilisin-type proteinase (EC 3.4.21.-) BH3763 precursor [similarity] - Bacillus halodurans

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: C84120

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: C84120

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-757 <STO>

A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA07482.1; GSPDB:CN00

A:Experimental source: strain C-125

C:Genetics:

C:Gene: BH3763

C:Superfamily: microbial serine proteinase vpr; subtilisin homology

C:Keywords: hydrolase; serine proteinase

F:1-20/Domain: signal sequence #status predicted <SIG>

Query Match 11.5%; Score 259.5; DB 2; Length 757;
Best Local Similarity 22.6%; Pred. No. 5.6e-09;
Matches 119; Conservative 58; Mismatches 155; Indels 195; Gaps 20;

QY 8 VKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRG-----KITALYALG 57

Db 119 VRGMLDEBEGVHLTGKVKVAVIDTGDITHPD--LOSSYKGGYDFVDYDDDPMTIASQG 176

QY 58 RTNNANDTNGHGHVAGSVLGGSTNKGMAQANLVFQSIIMDSGGGLGGLPSNLQTLFSQ 117

Db 177 -----PPTLHGTHVSGIITAANGV--KGVAPEAIIYAYRALPGG--QGTEQVIAAIEK 227

QY 118 AYSAGARIHTNSGWAACVAYTTDSRVDDYVRKNDMTILFAAGNECPNGGTTISAPGTAK 177

Db 228 AVEDGVDVNLGLNTVNGPDWPTSLALDAAEVGVAVT--SNGNSGPNMTYVSGPTSK 286

QY 178 NAITVGAT-----EN--LRPSFGS----- 194

Db 287 KAISVGASAPPLNTPLYTAFGENEISLFPFSGGLPWAFKRLDPMIDVGVTKEWEGVD 346

QY 195 -----YADNINH----- 201

Db 347 AEGKVLIKRGVMPTEKVMHAAKARGVLIYNNTPGPTMGIEGVNIPVVSITREDG 406

QY 202 -----VAQFSRGP--TKDCRIKPDVMACTFILSRSS 233

Db 407 EPLLQLEQLQKNKELTIRTYRKEDFVALFSSRGVPTHTWDVKPDVAPGVSI-----DS 462

QY 234 LAPDSFWANHDSKYAYMGSTMATPIVAGNVAOLREHFVKNRGITPK--PSLLKALIA 291

Db 463 TIPNNG-----YLGNGTSMAPHVAGAAALIKQ-----AHPWTPEQVKAALMN 507

QY 292 GAADI-----GLYPNGQGWGRVTLDKSLNVAVYVNESSLSSTQKATYSFTATAGPLKI 347

Db 508 TAKKLVDQEGVPHEITQAGAGRIQVDKAV-----AATSLVYPCALSEFGK---- 551

QY 348 SLVWSDAPASTTASVTL--VNLDLVLITAPNGTGYVNDFTSPYNDNW 393

Db 552 ---WSKDDLRKRPVTLTIENHDTV----KRTYHISPPFDVDPGVGV 591

RESULT 15

A35742

aqualysin (EC 3.4.21.-) I precursor - Thermus aquaticus

C:Species: Thermus aquaticus

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: A35742; S00620; S00324

R:Terada, I.; Kwon, S.T.; Miyata, Y.; Matsuzawa, H.; Ohta, T.

J. Biol. Chem. 265, 6576-6581, 1990

A:Title: Unique precursor structure of an extracellular protease, aqualysin I, with

A:Reference number: A35742; MUID:90216674; PMID:2182621

A:Accession: A35742

A:Molecule type: DNA

A:Residues: 1-513 <TER>

A:Cross-references: GB:J90108; GB:D90108; GB:J05414; NID:g217171; PIDN:BAAL4135.1;

R:Kwon, S.T.; Terada, I.; Matsuzawa, H.; Ohta, T.

Eur. J. Biochem. 173, 491-497, 1988

A:Title: Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline

A:Reference number: S00620; MUID:88225062; PMID:3286255

A:Accession: S00620

A:Molecule type: DNA

A:Residues: 75-442 <KWC>

A:Cross-references: EMBL:X07734; NID:g48069; PIDN:CAA30559.1; PID:g602091

A:Note: part of this sequence, including the amino and carboxyl ends of the mature

R:Matsuzawa, H.; Tokugawa, K.; Hamaoki, M.; Mizoguchi, M.; Taguchi, H.; Terada, I.;

Eur. J. Biochem. 171, 441-447, 1988

A:Title: Purification and characterization of aqualysin I (a thermophilic alkaline

A:Reference number: S00324; MUID:88151937; PMID:3162211

A:Accession: S00324

A:Molecule type: protein

A:Residues: 128-170 <MATS>

C:Superfamily: subtilisin; subtilisin homology

C:Keywords: extracellular protein; hydrolase; serine proteinase

F:1-14/Domain: signal sequence #status predicted <SIG>

F:15-127/Domain: propeptide #status predicted <PRO>

F:128-408/Product: aqualysin I #status experimental <MAT>

F:157-364/Domain: subtilisin homology <SBT>

F:255-257, 281-283/Region: S1 specificity crevice #status predicted

F:409-513/Domain: carboxyl-terminal propeptide #status predicted <CPR>

F:166,197,349/Active site: Asp, His, Ser #status predicted

Query Match 11.5%; Score 257.5; DB 1; Length 513;
Best Local Similarity 26.0%; Pred. No. 4.5e-09;
Matches 117; Conservative 44; Mismatches 144; Indels 145; Gaps 23;

QY 16 SYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALLY-ALGRTNNANDTNGHGHVAG 74

Db 152 TYTATGRGVNVVVDITGIRT-----THREFGGRARVGYDALG--GNGQDCNGHGHVAG 203

QY 75 SVLGNSTNKGMAQANLVFQSIIMD--SGGGLGGLPSNLQTLFESQAYSAGARIHTN---- 128

Db 204 TI---GGVTYGVAKAVNLYAVRVLDPCNGSGSTSGVIAGVDW-----TRNHRPVA 252

QY 129 --SWGAAVNGAYTTDSRVNVDYVRKN---DMTILFAAGNEGPNGGTTISAPGTAKNALTVG 183

Db 253 NWSLGGGVSTA-----LDNAVKNLSAAGVYVAVAGNDNANACNYS-PARVAEALTVG 304

QY 184 AT--ENLRPSFGSYADNINHVAFQSSRGPTKGRTPKDVYMAPGTFFLSARSSLAPOSSFW 241

Db 305 ATTSDDARASFNSYSGCV-----DLFAPGASIPSA-----W 335

QY 242 ANHLSKAYMGCTSNATPIVAGNVAOLREHFVKNRGITP---KPSLLKALIAAGADIGL 298

Db 336 YTSOTATOTLNGTSMATPHVAG--VAAL--YLEQNPSATPASVASAILNGATTGRLSGIGS 392

QY 299 GYPNGNQGWGRVTLDKSLNVAVYVNESSLSSTQKATYSFTATAGKPLKISLVWSDAPAST 358

Db 393 GSPN-----RLLYSLLSSGS-----GSTAPCTS 415

QY 359 ---TASVTLVNDLVLITAPNGTGY-----VGNDFTSYNDNDGRN--- 397

Db 416 CSYYTGSLSGPGDYNE---QPNGTYYYYSPAGTHRAWLRGPAGTDF-DLYLWRWDGSRWLT 471

Qy 398 -----NVENVFINAPQSGTYTIEVQAYN 420

Db 472 VGSSTGPTSEESLSYSGTAGYVYLWRTIYAYS 501

Search completed: July 25, 2003, 19:02:03
Job time : 20.0185 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:56:30 ; Search time 9.01038 seconds
(without alignments)
2265.120 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAOSSYGLY.....EVQAYNVPGPQTFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	523.5	23.3	1743	1	TAGC_DICDI
2	497	22.1	1905	1	TAGE_DICDI
3	310.5	13.8	806	1	SUBV_BACSU
4	304.5	13.6	1398	1	PLS_PYRFO
5	283	12.6	580	1	EXPR_XANCP
6	277	12.3	420	1	SUBT_BACSP
7	257.5	11.5	513	1	AQLI_THEAQ
8	253.5	11.3	894	1	WPRB_BACSU
9	246.5	11.0	402	1	ALP_CEPAC
10	246.5	11.0	534	1	PROA_VIBAL
11	242.5	10.8	401	1	THES_BACSP
12	240	10.7	269	1	SUBS_BACLE
13	240	10.7	380	1	ELVA_BACAO
14	240	10.7	380	1	PRTM_BACSP
15	239	10.6	269	1	ELVA_BACSP
16	235.5	10.5	382	1	SUBB_BACLE
17	235	10.5	378	1	ISP_BACCS
18	232	10.3	269	1	SUBT_BACLI
19	231	10.3	321	1	SUBT_BACLI
20	230.5	10.3	379	1	SMP1_MAGPO
21	230	10.2	404	1	SCA2_STRPY
22	230	10.2	1181	1	SCAL_STRPY
23	227	10.1	1167	1	SUBN_BACNA
24	225.5	10.0	381	1	MSIP_CRIGR
25	225.5	10.0	1052	1	MSIP_HUMAN
26	225.5	10.0	1052	1	MSIP_MOUSE
27	225.5	10.0	1052	1	MSIP_RAT
28	223.5	9.9	381	1	SUBT_BACSA
29	223.5	9.9	381	1	SUBT_BACST
30	223.5	9.9	381	1	SUBT_BACSU
31	223.5	9.9	381	1	SUBT_BACPU
32	221.5	9.9	275	1	HLX_HALL17
33	221.5	9.9	530	1	HLX_HALL17

ALIGNMENTS

RESULT 1
TAGC_DICDI

ID	TAGC_DICDI	STANDARD	PRT	1743 AA
AC	Q23868			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Prestalk-specific protein tagC precursor (EC 3.4.21.-)			
GN	TAGC			
OS	Dictyostelium discoideum (Slime mold)			
OC	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium			
OX	NCBI_TaxID=44689;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AX4			
RX	MEDLINE=97140317; PubMed=8986798;			
RA	Shaulsky G., Escalante R., Loomis W.F.;			
RT	"Developmental signal transduction pathways uncovered by genetic suppressors."			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).			
CC	FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGC MAY MEDIATE INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS (BY SIMILARITY).			
CC	SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY S8.			
CC	SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.			
CC	SIMILARITY: STRONG, TO TAGC.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U60086; AAB03331.1; -
PIR: T18279; T18279.
DictyDB: DD02060; tagC.
InterPro: IPR003593; AAA_Atpase.
InterPro: IPR001140; ABC_TM_transp.
InterPro: IPR0003439; ABC_transporter.
InterPro: IPR000209; Peptidase_S8.
Pfam: PF00664; ABC_membrane; 1.
Pfam: PF00005; ABC_tran; 1.
Pfam: PF00082; Peptidase_S8; 1.
PRINTS: PF00723; SUBTILISIN.
SMART: SM00382; AAA; 1.
PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
PROSITE: PS50893; ABC_TRANSPORTER_2; 1.
PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
PROSITE: PS00137; SUBTILASE_HIS; 1.
PROSITE: PS00138; SUBTILASE_SER; FALSE_NEG.
Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane; Signal.

P00781 bacillus li
P16396 bacillus su
P58502 pyrococcus
P41363 bacillus ha
P04072 thermocactin
P20015 tritirachiu
P29139 paenibacill
Q03420 trichoderma
P11018 bacillus su
P16397 bacillus su
P29138 metarhizium
P23653 tritirachiu

34 216.5 9.6 274 1 SUBD_BACLI
35 216.5 9.6 645 1 SUBE_BACSU
36 215 9.6 422 1 TKSU_PYRKO
37 214.5 9.5 361 1 ELYA_BACHD
38 214 9.5 279 1 THET_THEVU
39 214 9.5 293 1 PRPT_TRIAL
40 214 9.5 326 1 ISP_PAEPO
41 214 9.5 409 1 ALP_TRIHA
42 211 9.4 319 1 ISPI_BACSU
43 209.5 9.3 1433 1 SUBF_BACSU
44 208.5 9.3 388 1 CUDP_METAN
45 208 9.3 387 1 PRTR_TRIAL

FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 1743 PRESTALK-SPECIFIC PROTEIN TAGC.
 FT DOMAIN 316 642 PROTEASE.
 FT DOMAIN 1450 1687 ABC TRANSPORTER.
 FT TRANSMEM 962 982 POTENTIAL.
 FT TRANSMEM 1027 1047 POTENTIAL.
 FT TRANSMEM 1072 1092 POTENTIAL.
 FT TRANSMEM 1157 1177 POTENTIAL.
 FT TRANSMEM 1260 1280 POTENTIAL.
 FT TRANSMEM 1288 1308 POTENTIAL.
 FT ACT_SITE 325 325 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 372 372 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 637 637 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1485 1492 ATP (POTENTIAL).
 FT DOMAIN 42 46 POLY-ASN.
 FT DOMAIN 94 103 POLY-ASN.
 FT DOMAIN 643 646 POLY-ALA.
 FT DOMAIN 733 741 POLY-ASN.
 FT DOMAIN 786 792 POLY-SER.
 FT DOMAIN 1337 1340 POLY-GLY.
 FT DOMAIN 1346 1352 POLY-GLY.
 FT DOMAIN 1353 1357 POLY-ASN.
 FT DOMAIN 1358 1364 POLY-ASP.
 FT DOMAIN 1381 1386 POLY-ASN.
 FT DOMAIN 1707 1729 POLY-ASN.
 FT CARBOHYD 390 390 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 547 547 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 614 614 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 689 689 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 741 741 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 776 776 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 832 832 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 887 887 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1251 1251 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1385 1385 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1386 1386 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1454 1454 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1704 1704 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT SEQUENCE 1743 AA; 194145 MR; 12DB363E2F729839 CRC64;
 Query Match 23.3%; Score 523.5; DB 1; Length 1743;
 Best Local Similarity 27.9%; Pred. No. 2.4e-26;
 Matches 165; Conservative 81; Mismatches 150; Indels 195; Gaps 24;
 69 LYGGQGVAVADTGLDTCR---NDS-----SMHEAFRGKITALYALGRTNNANDTNGH 68
 314 LRKGQILLSIADTGLDGHCHCFSDSKYPIPLNSVNLNHR-KVVTYITTTSDSDSKVDGH 372
 69 GTHVAGSYLG-----NGSTNKGMAPQANLVFQSIMDSGGGLGL--PSNLOTLPFSQAY 119
 373 GTHICGSAAGTPEDSSVNISSFGLTDAKIAF---FDLASGSSSLTPPSDLKQLYOPLY 429
 120 SAGARIHTNSGA---AVNGAYTTDSNRVDVVRKN-DWTILFAAGNEPGNGTIS--A 172
 430 DAGARVHCDWSGVSVEGYTGSYSSTASIDDFLFTHPDFIILRAAGN---NEQYLSLLT 486
 173 PGTAKNATVGTATNLR-----PSFGSYADNI----- 199
 487 QSTAKNVITVGAHQTHIENYLTDPNPNYINQSSVDINQELICDFDSRYCNYTTAQCCLES 546
 200 -----NHVAOFSSRGPTKDKGRIPKDPVWAPGTFFIL 228
 547 NATTGLASCCPTLLRKSVIDAANTOPLLYNNENNCSFKGPTHDGRMKPALVAPGEYIT 606
 229 SARSSLA-----PDSSFWANHDSKYVYMGSTMATPIVAGNVAQLREH-----F 272
 607 SARNAGANTDOCGGSL-PNTNALLA-ISGTSMATSFARAAATILRQYLVGDVYPTGSI 664
 273 VKNRGITPKSLKKAALTAGA-----ADIGLYPNCN-----QGWGRVT 311
 665 VESNKLQPTGSLKALMINNAQLNCTFOLITSSSITYPNSQVFNFPAGASLVQGWGAIR 724

QY 312 LDKSLNAVYVNESS-----SLSTSQKATYSFT-- 338
 DB 725 MSNWLHVNNNNNNNNKTSIGITKFDGIGGLDLRLVKFNOWKEESLSTGQNTSYCYFTYK 784
 QY 339 -----ATAGKPLK---ISLWSDAPASTTASVTIYNDLVL-----TAPNGT 378
 DB 785 PSSSSSSNGNIPRVATLVWTDPPSYAGAKFNVLNLDLTMIYYRDNGSTIFYSNOGGS 844
 QY 379 QYGVNDFTSYNDANWGNVNFVINAPOSCTYTIYEQAYNVVPGPQTF 429
 DB 845 SFLG---LAPQDT---LNNVEGIVHNTEPTMYRFVWAGTVWPMGPQNF 889
 RESULT 2
 TAGB_DICDI STANDARD; PRT; 1905 AA.
 ID TAGB_DICDI
 AC P54683;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Prestalk-specific protein tagB precursor (EC 3.4.21.-).
 GN TAGB.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=AX4.
 RX MEDLINE=95262903; PubMed=7744252;
 RA Shaulsky G., Kuspa A., Loomis W.F.;
 RT "A multidrug resistance transporter/serine protease gene is required
 for prestalk specialization in Dictyostelium.";
 RL Genes Dev. 9:1111-1122(1995).
 CC -!- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
 CC -!- INTEGRATION OF CELLULAR DIFFERENTIATION WITH MORPHOGENESIS.
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
 CC S8.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
 CC -!- TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
 CC -!- SIMILARITY: STRONG, TO TAGC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U20432; AAA62212.1; -
 CC PIR; T18267; T18267.
 CC MEROPS; S08.0P.W; -
 CC DictyDb; DD02059; tagB.
 CC InterPro; IPR003593; AAA_ATPase.
 CC InterPro; IPR001140; ABC_TM_transport.
 CC InterPro; IPR003439; ABC_transporter.
 CC InterPro; IPR000209; Peptidase_S8.
 CC Pfam; PF00664; ABC_membrane; 1.
 CC Pfam; PF00005; ABC_tran; 1.
 CC Pfam; PF00082; Peptidase_S8; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC ProDom; PD000006; ABC_transporter; 1.
 CC SMART; SM00382; AAA; 1.
 CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
 CC PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
 CC PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
 CC PROSITE; PS00137; SUBTILASE_HIS; 1.
 CC PROSITE; PS00138; SUBTILASE_SER; 1.
 CC KW. Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
 KW Signal.
 FT SIGNAL. 1 31 POTENTIAL.
 FT CHAIN 32 1905 PRESTALK-SPECIFIC PROTEIN TAGB.
 FT

"The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*,"
 Nature 390:249-256(1997).
 CC -!- FUNCTION: NOT REQUIRED FOR GROWTH OR SPORULATION.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- PTM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M76590; AAA22881.1; -;
 CC EMBL; X73124; CAAS1601.1; -;
 CC EMBL; Z99123; CAB15835.1; -;
 CC PIR; A41341; A41341.
 CC HSP; P00782; 2SPT.
 CC MEROPS; S08.09A; -;
 CC Subtilisin; Bg10591; vpr.
 CC InterPro; IPR003137; PA.
 CC InterPro; IPR000209; Peptidase_S8.
 CC Pfam; PF02225; PA; 1.
 CC Pfam; PF00082; Peptidase_S8; 1.
 CC PRINTS; PR00723; SUBTILISIN.
 CC PROSITE; PS00136; SUBTILASIN.ASP; 1.
 CC PROSITE; PS00137; SUBTILASIN.HIS; 1.
 CC PROSITE; PS00138; SUBTILASIN.SER; 1.
 CC Hydrolyase; Serine protease; zymogen; Signal; Complete proteome.
 KW SIGNAL 1 28
 FT PROPEP 29 160
 FT CHAIN 161 806
 FT ACT_SITE 189 189
 FT ACT_SITE 233 233
 FT ACT_SITE 534 534
 FT ACT_SITE 806 806
 FT SEQUENCE 806 AA; 85608 MW; F984E3BF0B869DDD CRC64;
 Query Match 13.8%; Score 310.5; DB 1; Length 806;
 Best Local Similarity 23.5%; Pred. NO. 6.6e-13;
 Matches 129; Conservative 58; Mismatches 157; Indels 205; Gaps 18;
 QY 18 GLVGQGIIVADVGLDTGR-----NDSSMHEAFRGKITALYALGRTN 60
 DB 177 GTYGGKIGKVALIDRGVEYNHPDLKKNFGQYGYGVDFVNDYDKPTG-----D 225
 QY 61 NANTNGHGHVAGSVLGNSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 DB 226 PRGEATDGHVAGTVAANG-TIKGVAPDATTLLAYRVLPGG--SGTTENVIAQVERAVQ 282
 QY 121 AGARIHNSGAANGAYTTDSRNVDVVRKNDWTLILFAAGNEGNGGTISAPGTAKNAI 180
 DB 283 DGADVMNLSLGNLNPDPWATSTAL-DWAMSEGVAVTNSGNGPNGWTGSPGTSREAI 341
 QY 181 TVGATE---NLRFSGSY-----195
 DB 342 SVGATQLPLNEYATFGSYSSAKYMGYNKEDDVKALNNKEVELVEAGIGKDFGKDLT 401
 QY 196 -----ADNI-----199
 DB 402 GKVAVKRGSIATFVDKADNARKAKAIGMIVVYNNLSGEIEANVPMSVPTIKLSLEDGEKL 461
 QY 200 -----NHVAQSSRGPTKD-GRIKPDVMAAGFFILSARSLAP 236
 DB 462 VSALKAGETKTKTKLTVSKALGEQVADFSSRGVPMDWIKPDISAPGVNIVSTIPTHDP 521
 QY 237 DSSEWANHDSKYAYMGTSMAPTIVAGNVAQALREHFVKNRGITPKPSL--LKAALTAGAA 294
 DB 522 D-----HPYGYGSKQGTSMASPHIAGAVAIKQ-----AKPKWSVEQIKAAIMNTAV 568
 QY 295 DI-----GLGYPNGNOGWGRVTLDKSLNVAIVYVNESSLSSTSQKATYSFTAGPLKLSLV 350

Db 569 TLKDSGEYYPHNAQAGSARI-----MNAIKADSLVSPGSYSY-----CTFLK----- 612
 QY 351 WSDAPASTTASVTLVNDLDELVITAPNGTQYVGNDETSPYNDWDG--RNNVENVFNAQ 408
 Db 613 -ENGTNETFTIENG-----SSIRKSYTLEYSFNGSGISTSGTSRVVIPAHO 660
 QY 409 SGTYTIEVQ 417
 Db 661 TGRKATAVK 669
 RESULT 4
 PLS_PYRFU
 ID PLS_PYRFU STANDARD: PRT: 1398 AA.
 AC P72186;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pyrolysin precursor (EC 3.4.21.-).
 GN PLS OR PF0287.
 OS *Pyrococcus furiosus*.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC *Pyrococcus*.
 OC NCBI_TaxID=2261;
 RN [1]
 RC SEQUENCE FROM N.A., SEQUENCE OF 150-184, AND CHARACTERIZATION.
 RP STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
 RX MEDLINE=96355370; PubMed=8702780;
 RA Voorhorst W.G.B., Eggen R.I.L., Geerling A.C.M., Platteuw C.,
 RA Siezen R.J., de Vos W.M.;
 RA "Isolation and characterization of the hyperthermostable serine
 RT protease, pyrolysin, and its gene from the hyperthermophilic archaeon
 RT *Pyrococcus furiosus*,"
 RT J. Biol. Chem. 271:20426-20431(1996).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
 RX Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RA "The complete sequence of the *Pyrococcus furiosus* genome,"
 RT Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RN CHARACTERIZATION, AND 3D-STRUCTURE MODELING.
 RP MEDLINE=21079021; PubMed=11210516;
 RX de Vos W.M., Voorhorst W.G.B., Dijkgraaf M., Kluskens L.D.,
 RA van der Oost J., Siezen R.J.;
 RA "Purification, characterization, and molecular modeling of pyrolysin
 RT and other extracellular thermostable serine proteases from
 RT hyperthermophilic microorganisms,"
 RL Meth. Enzymol. 330:383-393(2001).
 CC -!- FUNCTION: Has endopeptidase activity toward caseins, casein
 CC fragments including alpha-S1-casein and synthetic peptides.
 CC -!- SUBCELLULAR LOCATION: Cell-envelope associated.
 CC -!- PTM: LWM pyrolysin seems to be produced by autoprolytic
 CC activation of HWV pyrolysin.
 CC -!- PTM: Glycosylated.
 CC -!- MISCELLANEOUS: Thermostable; high activity at 95 degrees Celsius.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U55835; AAB09761.1; -;
 CC EMBL; AE010153; AAL80411.1; -;
 CC PIR; T28159; T28159.
 CC HSP; Q45670; LDB1.
 CC MEROPS; S08.100; -;
 CC InterPro; IPR000209; Peptidase_S8.


```

DR Pfam: PF00082; Peptidase_S8; 1.
DR Pfam: PF04151; PPC; 1.
DR DR PROSITE: PR00723; SUBTILISIN.
DR DR PROSITE: PS00136; SUBTILASE ASP; 1.
DR DR PROSITE: PS00137; SUBTILASE HIS; 1.
DR DR PROSITE: PS00138; SUBTILASE SER; 1.
DR KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL 1 ? POTENTIAL.
FT PROPEP 149 ?
FT CHAIN 150 1398 PYROLYSIN.
FT ACT_SITE 179 179 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 365 365 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 152 152 N-LINKED (PROBABLE).
FT CARBOHYD 222 222 N-LINKED (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (POTENTIAL).
FT CARBOHYD 240 240 N-LINKED (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (POTENTIAL).
FT CARBOHYD 327 327 N-LINKED (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (POTENTIAL).
FT CARBOHYD 663 663 N-LINKED (POTENTIAL).
FT CARBOHYD 739 739 N-LINKED (POTENTIAL).
FT CARBOHYD 792 792 N-LINKED (POTENTIAL).
FT CARBOHYD 893 893 N-LINKED (POTENTIAL).
FT CARBOHYD 908 908 N-LINKED (POTENTIAL).
FT CARBOHYD 917 917 N-LINKED (POTENTIAL).
FT CARBOHYD 929 929 N-LINKED (POTENTIAL).
FT CARBOHYD 1048 1048 N-LINKED (POTENTIAL).
FT CARBOHYD 1056 1056 N-LINKED (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (POTENTIAL).
FT CARBOHYD 1117 1117 N-LINKED (POTENTIAL).
FT CARBOHYD 1133 1133 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1148 1148 N-LINKED (POTENTIAL).
FT CARBOHYD 1208 1208 N-LINKED (POTENTIAL).
FT CARBOHYD 1233 1233 N-LINKED (POTENTIAL).
FT CARBOHYD 1237 1237 N-LINKED (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (POTENTIAL).
FT CONFLICT 607 609 AKA -> PKP (IN REF. 1).
FT CONFLICT 881 881 Y -> H (IN REF. 1).
SQ SEQUENCE 1398 AA; 154474 MW; 355D873A27D56552 CRC64;

Query Match
Best Local Similarity 26.6%; Score 304.5; DB 1; Length 1398;
Matches 139; Conservative 57; Mismatches 178; Indels 149; Gaps 22;

QY 21 GGGIVAVADTGLDGTGRNDS-----SMHFAFRGKITALYALGRNAN----- 63
DB 301 GNGYDAYVDLDDYDFTDEVLPGQVNTYDVAVFYSYPLNVLAEIDPNGEYAVFGW 360
QY 64 DTNGHGHVAGSVLNGSTN-----KGMAPQAN 91
DB 361 DGHGHGHVAGVAGVYDSNDANDWLSMYSGEWEFSRLYGWDTYNTVTTVOGVAPGAQ 420
QY 92 LVFQSIMDSGGGLGSLNQLTFLSQAYSAGARIHTNSMGAAYNGAYT--TDSRN--VDD 147
DB 421 IMAIRVLRS-DGRGSMWDIEGM-TYAATHGADVISMSLGG--NAPYLDGTDPEVAVDE 476
QY 148 YVRKNDMTLFAAGNGPNNGGTISAPCTAKNAITVCATENLRFSGSYAD----- 197
DB 477 LTEKGVVVFVIAAGNEGPGINIVGSPVATKAITVGAAA-VPINVGVIYSQALGYDYYG 535
QY 198 -----NINHAOFSSRGPKDKGRIPKDVMAPGTFILSARSSLAPDSSFANHSKYA 249
DB 536 FYFFPAYTNV-RIAFESSRGRIDGELKPNVAVPGYIYSSLPWLGADF----- 585
QY 250 YMGGTSMATPIVAGNVVQALREHFVKNRGTTPKPSLLKAALIAAGADI-----GLGYPNG 303
DB 586 -MSGTSMATPHVSGYVALLISG-AKAEGIYYPNDIHKVLESATWLEGDPYTGQYKYTEL 643

```

```

QY 304 NOGWRVTLQKSLNAVYVYNESSLSSTQKATYSFTATAGKPLKISLWNSDAPASTTA--- 360
DB 644 DQGHGLNVNTKSWEI-----LKAINGTTLPIVDHWADKSYSDFAEYL 685
QY 361 SVTLVNDLDLVTAPN-----GTQVGN-----DFTSPYNDNW-----DG-----RNNVENVF 403
DB 686 GVDVIRGLYARNSIPDIVWEHIKYVGDTEYRTFEIYATPEWIKPFVSGSVILENTEFVL 745
QY 404 -----INAPSGTY-----TIEVQAYNVPVGPOTFS 429
DB 746 RVKYDVEGLEPGLVVGRIIDDPPTPVIIEILNITVIVPEKPT 788

RESULT 5
ID EXPR_XANCP STANDARD; PRT; 580 AA.
AC P2314;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Extracellular protease precursor (EC 3.4.21.-).
GN XCC0851.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_taxid=340;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90251253; PubMed=2187155;
RT Liu Y.-N., Tang J.-L., Clarke B.R., Dow J.M., Daniels M.J.;
RT "A multipurpose broad host range cloning vector and its use to
RT characterise an extracellular protease gene of Xanthomonas campestris
RT pathovar campestris.";
RL Mol. Gen. Genet. 220:433-440(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferri J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferri M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to peptidase family S8.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X51635; CAA35962.1; -.
DR EMBL; AE012184; AAM40166.1; -.
DR PIR; S11890; S11890.
DR HSP; P00782; 2SPT.
DR MEROPS; S08.0PA; -.

```

DR	InterPro: IPR000209; Peptidase_S8.	RP	SEQUENCE FROM N.A.
DR	Pfam: PF00082; Peptidase_S8; 1.	RX	MEDLINE-92256481; PubMed=1581352;
DR	PRINTS: PR00723; SUBTILISIN.	RT	Narinx E., Davail S., Feller G., Gerday C.;
DR	PROSITE: PS00136; SUBTILASE-ASP; 1.	RA	"Nucleotide and derived amino acid sequence of the subtilisin from
DR	PROSITE: PS00137; SUBTILASE-HIS; 1.	RL	the antarctic psychrotroph <i>Bacillus</i> TA39.";
DR	PROSITE: PS00138; SUBTILASE-SER; 1.	CC	Biochim. Biophys. Acta 1131:111-113(1992).
KW	Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.	CC	-!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
FT	SIGNAL 1 32	CC	-!- IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
FT	POTENTIAL.	CC	-!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
FT	PROPEP 33 7136	CC	for peptide bonds, and a preference for a large uncharged residue
FT	CHAIN 2137 580	CC	in <i>Pl.</i> Hydrolyzes peptide amides.
FT	ACT_SITE 177 177	CC	-!- SUBCELLULAR LOCATION: Secreted.
FT	ACT_SITE 237 237	CC	-!- MISCELLANEOUS: STILL ACTIVE AT TEMPERATURES CLOSE TO 0 DEGREES
FT	ACT_SITE 409 409	CC	CELSIUS, IT HAS A MARKED HEAT STABILITY.
FT	DISULFID 225 273	CC	-!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
FT	DISULFID 315 352	CC	SPORULATION AND MANY MUTATIONS WHICH BLOCK SPORULATION AT EARLY
FT	DISULFID 450 454	CC	STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
FT	POTENTIAL.	CC	IS NOT NECESSARY FOR NORMAL SPORULATION.
SEQUENCE	580 AA; 57228 MW; 8C9A2CAB4E7F47CB CRC64;	CC	-!- SIMILARITY: Belongs to peptidase family S8.

Query Match	12.6%; Score 283; DB 1; Length 580;	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
Best Local Similarity	26.2%; Pred. No. 2.7e-11;	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
Matches 127; Conservative	55; Mismatches 147; Indels 156; Gaps 24;	CC	the European Bioinformatics Institute. There are no restrictions on its
QY	21 GGGQIVAVADTGL-----DTGRNDSSMHEAFRGKITALYALGRITNNAND-----64	CC	use by non-profit institutions as long as its content is in no way
Db	168 GSGTVVAVIDGITSHADLNAILAGYDFISDATTARDGCRDSNADEGDWYAANECCA 227	CC	modified and this statement is not removed. Usage by and for commercial
QY	65 -----TNGHGTHTVAGSVLNGSTNKGMAPOA-----NLVFQSI MD 99	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
Db	228 GIPAASSSWHGTHTVAGVAAVTNNTTGVAGTAYGAKVVPVRLGKCGSLSDIADAIWVA 287	CC	EMBL: X62369; CAA44227.1; -
QY	100 SGGGLGLPSNLQ--TLFSQAYSAGARHTHTNSGAAVNGAVYTTDSRNVDVYVRKNDMTIL 157	DR	PIR: S23407; S23407.
Db	288 SGTGVSIPANANPAEVIINSLGGGCGSSTTMQN-AINGAVSRGT-----TVV 334	DR	HSP: Q99405; IMPT.
QY	158 FAAGNEGPGGTISAPGTAKNAITVGATEN--LRPFGSYADNINHVAFQSSRGPTKQGR 215	DR	MEROPS: S08.0PA; -
Db	335 VAAGNDASNSVG--SLPANCANVIAVAATTSAGAKASYNFGGI-----377	DR	InterPro: IPR000209; Peptidase_S8.
QY	216 IKPDVMAPGTFILSARSS--LAPDSSFWANHDSKYAYWGTSMAPIVAGNVAQLREHFV 273	DR	Pfam: PF00082; Peptidase_S8; 1.
Db	378 ---DVSAPGSLSLTSNGSTTTPGS-----ASYASNGTSMASPHVAGVVALVQS--V 425	DR	PRINTS: PR00723; SUBTILISIN.
QY	274 KNRGITPK--PSLLK--AALIAGAADIIGLYPNGNGWGRVTLDRKSLNVAIVYNESS-----325	DR	PROSITE: PS00136; SUBTILASE-ASP; 1.
Db	426 APTALTPAAVETLLKNTARALPGAC-----SGCGGAGIVNADAATAA--INGSGGGG 477	DR	PROSITE: PS00137; SUBTILASE-HIS; 1.
QY	326 -----SLSTSOKATYSFTATACKPLKISLVNSDAPASTTASVTL---VND 367	DR	PROSITE: PS00138; SUBTILASE-SER; 1.
Db	478 GGGNTLTNGTPTVGLGAATGAELNYTITVPAG-----SGTLTVTTSGGSGD 523	KW	Hydrolase; Sporulation; Serine protease; Zymogen; Signal.
QY	368 LDLVI---TAPNGTOYVGNDFTSYPYNDWGRNNVENVFINAPQSGTYTIEVOAYNVPVG 424	FT	SIGNAL 1 26
Db	524 ADLYVRAGSAPTDSAYT---CRPYRS-----GNAETCTITAP-SGTYYVRLKAYS----569	FT	POTENTIAL.
QY	425 PQTFE 429	FT	PROPEP 27 111
Db	570 --TFS 572	FT	CHAIN 112 420

RESULT 6		FT	ACT_SITE 145 145
ID	SUBT_BACS9	FT	ACT_SITE 182 182
AC	P28842;	FT	ACT_SITE 360 360
DT	01-DEC-1992 (Rel. 24, Created)	SEQUENCE	420 AA; 44086 MW; AE4P121BD32B26EC CRC64;
DT	01-DEC-1992 (Rel. 24, Last sequence update)	Query Match	12.3%; Score 277; DB 1; Length 420;
DT	28-FEB-2003 (Rel. 41, Last annotation update)	Best Local Similarity	31.2%; Pred. No. 4.4e-11;
DE	Subtilisin precursor (EC 3.4.21.62).	Matches 98; Conservative	44; Mismatches 120; Indels 52; Gaps 16;
GN	SUBI.	QY	21 GGGQIVAVADTGLDTRNDSSMHEAFRGKITAL--YALGRIT---NNANDTNGHGTHTVAGS 75
OS	<i>Bacillus</i> sp. (strain TA39).	Db	136 GGGINIAVLDTGVNTN-----HPDLRNNVVEOCKDFVTGYTYTNNSCTDROGHGTHVAGS 189
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; <i>Bacillus</i> .	QY	76 VLNGSTNK---GMAPOANL-VFOSIMDSGGGLG-GLPSNLQTLFSQAYSAGARHTTN-S 129
OX	NCBI_TaxID=29336;	Db	190 ALADGGTNGVYGVPADLDLWAYKVLGDCSGVADDDIAAIAIRHAGDOATALTNTKVINMS 249
RN	[1]	QY	130 WGAAYNGAYTTDSRNVDVYVRKNDMTILFAAGNEGPGNGGTISAPGTAKNAITVGATENLR 189
		Db	250 LGSSGESSLITNAVN---YSYNGKVLIIAAAGNSPGYQGSITGYPCALYNVAVAALEN-K 305
		QY	190 PSFGSYADNINHVAFQSSRGPT-KDG-----RIKPDYMAPCTEILSARSLAPDSSFWAN 243
		Db	306 VENGTY-----RYADFSSRGYSWTDGDYATOKGDVEISAPGAAYST-----W---348
		QY	244 HDSKYAYMCGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALITAGAADIIGLYPNG 303
		Db	349 FDGGYATISGTSWASPHAAGLAAKIWAQYFSASNVDRGELQYRAY---ENDILSGYVAG 405
		QY	304 -----NQGWGRVTL 312

van der Laan J.C., Teplyakov A.V., Kelders H., Kalk K.H., Missset O., Mulleners L.J.M., Dijkstra B.W.;
 "Crystal structure of the high-alkaline serine protease PB92 from Bacillus alcalophilus.";
 Protein Eng. 5:405-411(1992).
 [3]
 RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
 RA MEDLINE=93078250; PubMed=1447775;
 RX Sobek H., Hecht H.-J., Ahle W., Schomburg D.;
 "X-ray structure determination and comparison of two crystal forms of a variant (Asn15Arg) of the alkaline protease from Bacillus alcalophilus refined at 1.85-A resolution.";
 J. Mol. Biol. 228:108-117(1992).
 [4]
 RN STRUCTURE BY NMR OF 112-380.
 RP STRAIN=PB92;
 RC MEDLINE=97277237; PubMed=9115441;
 RX Martin J.R., Mulder F.A., Karimi-Nejad Y., van der Zwan J., Mariani M., Schipper D., Boelens R.;
 "The solution structure of serine protease PB92 from Bacillus alcalophilus presents a rigid fold with a flexible substrate-binding site.";
 Structure 5:521-532(1997).
 RL -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M65086; AAA22212.1; -;
 DR EMBL: A13738; CAA01128.1; -;
 DR PIR: A49778; A49778;
 DR PDB: 1AH2; 15-APR-98.
 DR MEROPS: S08.038; -;
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PROSITE: PS00136; SUBTILASE_SP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 DR Hydrolase; Serine protease; zymogen; Signal; 3D-structure.
 FT SIGNAL 1 27
 FT PROPEP 28 112
 FT CHAIN 113 380
 FT ACT_SITE 143 143
 FT ACT_SITE 173 173
 FT ACT_SITE 326 326
 FT HELIX 117 120
 FT TURN 121 123
 FT TURN 125 130
 FT TURN 131 131
 FT TURN 135 136
 FT STRAND 138 142
 FT TURN 150 151
 FT STRAND 156 157
 FT HELIX 173 182
 FT STRAND 198 201
 FT TURN 207 208
 FT HELIX 213 226
 FT STRAND 230 233
 FT HELIX 242 252
 FT TURN 253 255
 FT STRAND 257 268
 FT TURN 272 275
 FT TURN 279 279
 FT STRAND 282 285
 FT STRAND 291 291

FT TURN 299 300
 FT STRAND 303 306
 FT TURN 310 314
 FT TURN 315 317
 FT STRAND 318 322
 FT HELIX 325 342
 FT TURN 344 345
 FT HELIX 348 358
 FT HELIX 375 380
 SQ SEQUENCE 380 AA; 539EA72771B6682C CRC64;
 Query Match 10.7%; Score 240; DB 1; Length 380;
 Best Local Similarity 31.1%; Pred. No. 9.8e-09;
 Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;
 QY 8 VKADVAQSSYGLYGGQIVAVADTGRNDSSMHEAFRGKITALYALGRTNANDNG 67
 DB 122 VQAPAAHNR-GLTSGVKVAVLDTGIST-----HPDINIRGGASFVGEPE-STODGNG 172
 QY 68 HGTHVAGSV--LNGSTNKGMAPQANLVFQSIMDSGGGLGLPSNLQTLFSQAYSAGARI 125
 DB 173 HGTHVAGTIAALNNSIGVLGVAPNAELYAVKVLGASG--SGSVSSIAOGLWAGNNGMHV 230
 QY 126 HTNSWGA-----AVNGAYTTDSNRVDDVVRKNDMTILFAAGNEGPNGGTISAPGTA 176
 DB 231 ANLSLGSFSPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 276
 QY 177 KNATTVGATE--NLRPSPFGSYADNINHVAFQSSRGPTKDRIKPDVYMAPGTFTLSARSSL 234
 DB 277 ANAMAVGATDNNNRASFQVAGL-----DIVAPGVNVQSTVPG- 316
 QY 235 APDSSFANHDSKYAYMGTSMTATPIVAGNVA-----QLREHFVKV 275
 DB 317 -----STVASLNGTSMATPHVAGAAALVKQKPNPSWNVQIRNH-LKN 357
 RESULT 14
 ID ELYA_BACCS STANDARD; PRT; 380 AA.
 AC P41362;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alkaline protease precursor (EC 3.4.21.-).
 OS Bacillus clausii.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=79880;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RX MEDLINE=93043753; PubMed=1368952;
 RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S., Aono R., Horikoshi K.;
 "Molecular cloning, nucleotide sequence, and expression of the structural gene for alkaline serine protease from alkaliphilic Bacillus sp. 221.";
 J. Bacteriol. 173:111-115(1991).
 [2]
 RP SEQUENCE OF 112-129.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RA Horikoshi K.;
 RL Microorganisms in alkaline environments, pp.187-194, VCH, Weinheim (1991).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC


```

CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; S48754; AAC60420.1; -
DR      EMBL; D13157; BAA02442.1; -
DR      EMBL; A26817; CAA01836.1; -
DR      EMBL; A22550; CAA01611.1; -
DR      HSSP; P29600; IGC1.
DR      MEROPS; S08.103; -
DR      InterPro; IPR000209; Peptidase_S8.
DR      Pfam; PF00082; Peptidase_S8; 1.
DR      PRINTS; PR00723; SUBTILISIN.
DR      PROSITE; PS00136; SUBTILASE_ASP; 1.
DR      PROSITE; PS00137; SUBTILASE_HIS; 1.
DR      PROSITE; PS00138; SUBTILASE_SER; 1.
KW      Hydrolase; Serine protease; Zymogen; Signal.
FT      SIGNAL      1 27
FT      PROPEP      28 111
FT      CHAIN      112 380
FT      ACT_SITE   143 143
FT      ACT_SITE   173 173
FT      ACT_SITE   326 326
SQ      SEQUENCE   380 AA; 38826 MW; 5F73ABC68D5B6831 CRC64;

Query Match      10.7%; Score 240; DB 1; Length 380;
Best Local Similarity 31.1%; Pred. No. 9.8e-09;
Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;

QY      8 VKADVAOSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNNANDTNG 67
      122 VQAPAAHNR-GLTSGGVKVAVLDTGIST-----HPDLNIRGGASFVPGEP-STQDNG 172
      68 HGTHVAGSV---LNGSTNKGMAPQANLVFQSIMDSGGGLGSLPSNLQTLFSAQYSAGARI 125
      173 HGTHVAGTIAALNNSIGVLGVAPSAELVAVKVLGASG--SGSVSSIAQGLEWAGNNGMHV 230
      126 HTNSWGA-----AVNGAYTTDSNRVDDYVRKNDMTTLFAAGNEGPNGTISAPCTA 176
      231 ANLSLGSPTSATLEQAVNSA---TSRGV-----LVAAAGSNGS--AGTSYPARY 276
      177 KNAITVGATE--NLRFSFGSYADINIHVAOFSSRPTKDGRIKPDVMAPGTFFLSARSSL 234
      277 ANAMAVGATDQNNRASFSYGAGL-----DIVAPGVNVQSTYFG- 316
      235 APDSSFWANHDSKYAWMGTSMATPIVAGNVA-----QUREHFVKN 275
      317 -----STYASLNGTSMATPHVAGAAALVKOKNPSWSNVQIRNH-LKN 357

RESULT 15
PRTM_BACSP STANDARD; PRT; 269 AA.
AC Q99405;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE M-protease (EC 3.4.21.-).
OS Bacillus sp. (strain KSM-K16).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RA Yamane T., Kani T., Hatanaka T., Suzuki A., Ashida T.,
RA Kobayashi T., Ito S., Yamashita O.;
RT "Structure of a new alkaline serine protease (M-protease) from
RT Bacillus sp. KSM-K16."
RL Acta Crystallogr. D 51:199-206(1995).
RN [2]
RP SEQUENCE OF 1-23, AND CHARACTERIZATION.
RX MEDLINE=95358832; PubMed=7632397;
RA Kobayashi T., Hakamada Y., Adachi S., Hitomi J., Yoshimatsu T.,
RA Koike K., Kawai S., Ito S.;
RT "Purification and properties of an alkaline protease from
RT alkalophilic Bacillus sp. KSM-K16."

```

```

RL      Appl. Microbiol. Biotechnol. 43:473-481(1995).
CC      -1- SUBUNIT: Monomer.
CC      -1- SIMILARITY: Belongs to peptidase family S8.
DR      PDB; IMPT; 22-JUN-94.
DR      MEROPS; S08.010; -
DR      InterPro; IPR000209; Peptidase_S8.
DR      Pfam; PF00082; Peptidase_S8; 1.
DR      PRINTS; PR00723; SUBTILISIN.
DR      PROSITE; PS00136; SUBTILASE_ASP; 1.
DR      PROSITE; PS00137; SUBTILASE_HIS; 1.
DR      PROSITE; PS00138; SUBTILASE_SER; 1.
KW      Hydrolase; Serine protease; 3D-structure.
FT      ACT_SITE   32 32
FT      ACT_SITE   62 62
FT      ACT_SITE   215 215
FT      HELIX      6 10
FT      TURN      11 12
FT      HELIX      13 18
FT      TURN      19 20
FT      TURN      24 25
FT      STRAND     27 32
FT      TURN      39 40
FT      STRAND     43 48
FT      TURN      51 52
FT      HELIX      62 71
FT      TURN      84 85
FT      STRAND     87 92
FT      TURN      96 97
FT      HELIX      102 114
FT      TURN      115 116
FT      STRAND     119 122
FT      STRAND     126 126
FT      HELIX      131 142
FT      TURN      143 144
FT      STRAND     146 150
FT      TURN      161 161
FT      TURN      162 164
FT      TURN      166 167
FT      STRAND     169 174
FT      TURN      176 177
FT      STRAND     180 180
FT      TURN      182 183
FT      TURN      188 189
FT      STRAND     192 195
FT      STRAND     199 203
FT      TURN      204 206
FT      STRAND     207 211
FT      HELIX      214 231
FT      TURN      233 234
FT      HELIX      237 246
FT      TURN      247 247
FT      STRAND     249 249
FT      HELIX      254 257
FT      TURN      258 259
FT      STRAND     261 261
FT      HELIX      264 267
FT      TURN      268 269
SQ      SEQUENCE   269 AA; 26723 MW; 7A03C86D534A1D07 CRC64;

Query Match      10.6%; Score 239; DB 1; Length 269;
Best Local Similarity 31.1%; Pred. No. 7.3e-09;
Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;

QY      8 VKADVAOSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTNNANDTNG 67
      11 VQAPAAHNR-GLTSGGVKVAVLDTGIST-----HPDLNIRGGASFVPGEP-STQDNG 61
      68 HGTHVAGSV---LNGSTNKGMAPQANLVFQSIMDSGGGLGSLPSNLQTLFSAQYSAGARI 125
      62 HGTHVAGTIAALNNSIGVLGVAPSAELVAVKVLGASG--SGSVSSIAQGLEWAGNNGMHV 119
      126 HTNSWGA-----AVNGAYTTDSNRVDDYVRKNDMTTLFAAGNEGPNGTISAPCTA 176

```

Tue Jul 29 14:12:43 2003

120	ANLSLGSPSPATLEQAVNSA	-----TSRGV-----	-----LVVAASGNSG-----	AGSTISYPARY	163
177	KNATIVGATE--NLRFSGSYADINIHVAOFSRSGPTKDGRIKDDVMAPGTFILSARSSL	234			
166	ANAMAYGATDQNNNRASFQYAGL	-----DIVAPGVNVOSTTPG--	205		
235	APDSSFWANHDSKIAYMGGTSMATPIVAGNVA	-----QLREHFVKV	275		
206	-----STYASLNGTSMATPHVAGVAALVKQKNPSWNSQVIRNH--LKN	246			

Search completed: July 25, 2003, 19:00:05
Job time : 12.0104 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:57:15 ; Search time 35.0404 seconds
(without alignments)
3196.166 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EQAYNPVGPQTFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_ivirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	640	2 Q93UV9	Q93UV9 bacillus sp
2	2183	97.2	639	2 Q9AQR3	Q9AQR3 bacillus sp
3	2143	95.4	434	2 Q9AQR0	Q9AQR0 bacillus sp
4	1998.5	88.9	433	2 Q9AQR1	Q9AQR1 bacillus sp
5	1994.5	88.8	433	2 Q9AQR4	Q9AQR4 bacillus sp
6	1987.5	88.5	433	2 Q9AQR2	Q9AQR2 bacillus sp
7	511.5	22.8	1825	5 Q8T9W1	Q8T9W1 dictyostell
8	447	19.9	1702	5 Q9GTN7	Q9GTN7 dictyostell
9	414	18.4	654	17 Q8U0C9	Q8U0C9 pyrococcus
10	406.5	18.1	561	16 Q8RBJ2	Q8RBJ2 thermococcus
11	398	17.7	1239	16 Q9FBZ4	Q9FBZ4 streptomyces
12	381	17.0	430	16 Q8ENV1	Q8ENV1 streptomyces
13	376	16.7	1253	16 Q9FC06	Q9FC06 oceanobacil
14	366	16.3	1102	2 P95684	P95684 streptomyces
15	349	15.5	1245	16 Q9RL54	Q9RL54 streptomyces
16	347.5	15.5	444	16 Q9KBJ7	Q9KBJ7 bacillus ha

17	346	15.4	1105	2 Q8KKH6	Q8KKH6 streptomyces
18	338	15.0	412	2 Q9AER6	Q9AER6 thermococcus
19	338	15.0	1237	2 Q8GGT4	Q8GGT4 streptomyces
20	337	15.0	412	16 Q8RC68	Q8RC68 thermococcus
21	328.5	14.6	1220	16 Q9LOA0	Q9LOA0 streptomyces
22	327.5	14.6	435	16 Q8EMJ3	Q8EMJ3 oceanobacil
23	324.5	14.4	824	2 Q45464	Q45464 bacillus sp
24	320.5	14.3	442	16 Q31788	Q31788 bacillus su
25	320.5	14.3	891	1 Q93635	Q93635 thermococcus
26	308.5	13.7	431	2 Q9P9L6	Q9P9L6 bacillus su
27	306.5	13.6	1398	1 Q9P9L1	Q9P9L1 pyrococcus
28	297	13.2	1135	1 Q9P9D1	Q9P9D1 uncultured
29	294.5	13.1	434	2 Q54327	Q54327 bacillus sp
30	282.5	12.6	586	16 Q8PAL8	Q8PAL8 xanthomonas
31	281	12.5	419	2 Q45681	Q45681 bacillus su
32	280	12.5	799	16 Q9KEM1	Q9KEM1 bacillus ha
33	279.5	12.4	1345	1 Q54437	Q54437 staphyloche
34	278.5	12.4	959	16 Q8PMS7	Q8PMS7 xanthomonas
35	277.5	12.3	621	2 Q9F486	Q9F486 alteromonas
36	277.5	12.3	621	2 Q53401	Q53401 alteromonas
37	272.5	12.1	568	16 Q8PMC0	Q8PMC0 xanthomonas
38	272.5	12.1	575	16 Q8PNW1	Q8PNW1 xanthomonas
39	268.5	11.9	403	2 Q45463	Q45463 bacillus sp
40	267	11.9	715	2 P70765	P70765 alteromonas
41	265	11.8	1098	16 Q9L128	Q9L128 streptomyces
42	264.5	11.8	966	16 Q8PB28	Q8PB28 xanthomonas
43	263	11.7	1331	17 Q9YEG9	Q9YEG9 aeropyrum p
44	263	11.7	1571	2 Q8GCW3	Q8GCW3 streptococ
45	262.5	11.7	467	2 Q93IQ4	Q93IQ4 xanthomonas

ALIGNMENTS

RESULT 1

Q93UV9 ID Q93UV9 PRELIMINARY; PRT; 640 AA.

AC Q93UV9; TREMBLrel. 19, Created

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Protease.

GN PROF.

OS Bacillus sp. KSM-KP43.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=109322;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-KP43;

RA Itoh S., Saeki K.

RT "new protease"

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB051423; BAB55674.2; -

DR InterPro; IPR000209; Peptidase_S8.

DR Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

SQ SEQUENCE 640 AA; 67991 MW; 4BBAF7E9D592C15 CRC64;

Query Match 100.0%; Score 2247; DB 2; Length 640;

Best Local Similarity 100.0%; Pred. No. 2, 4e-123;

Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYCGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60

Db 207 NDVARGIVKADVAQSSYGLYCGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266

QY 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLTQFLSQAYS 120

Db 267 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIIMDSGGGLGGLPSNLTQFLSQAYS 326

QY 121 AGARIHTNSGAAVNGAYTTTDSNRVDVYVRKNDMTILFAAGNEGPNCGTISAPGTAKNAI 180

Db	327	AGARIHNTNSWGAAYNGVATTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI	386	Db	386	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDCRIKPDVMAQGTIILSARSSLPDSSF	445
Qy	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSSLPDSSF	240	Qy	241	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY	300
Db	387	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSSLPDSSF	446	Db	446	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY	505
Qy	241	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY	300	Qy	301	PNGNOGWGRVTLDKSLNVAYNVNESSLSSTSQATYSFTATAGPKLISLVMSDAPASTTA	360
Db	447	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLGY	506	Db	506	PNGNOGWGRVTLDKSLNVAYNVNESSLSSTSQATYSFTATAGPKLISLVMSDAPASTTA	565
Qy	301	PNGNOGWGRVTLDKSLNVAYNVNESSLSSTSQATYSFTATAGPKLISLVMSDAPASTTA	360	Qy	361	SVTLVNDLDLVITAPNGTQYVGNDETSPYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN	420
Db	507	PNGNOGWGRVTLDKSLNVAYNVNESSLSSTSQATYSFTATAGPKLISLVMSDAPASTTA	566	Db	566	SVTLVNDLDLVITAPNGTQYVGNDETSPYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN	625
Qy	361	SVTLVNDLDLVITAPNGTQYVGNDETSPYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN	420	Qy	421	VPVGPQTFSLAIVN 434	
Db	567	SVTLVNDLDLVITAPNGTQYVGNDETSPYNDNWDGRNNVENVFINAPQSGTYTIEVQAYN	626	Db	626	VPVGPQTFSLAIVN 639	
Qy	421	VPVGPQTFSLAIVN 434		Qy	421	VPVGPQTFSLAIVN 434	
Db	627	VPVGPQTFSLAIVN 640		Db	626	VPVGPQNFSLAIVN 639	
RESULT 2							
Q9AQR3							
ID	Q9AQR3	PRELIMINARY;	PRT;	639	AA.		
AC	Q9AQR3;						
DT	01-JUN-2001 (TReMBLrel. 17, Created)						
DT	01-OCT-2002 (TReMBLrel. 22, Last sequence update)						
DT	01-MAR-2003 (TReMBLrel. 23, Last annotation update)						
DE	Protease.						
GN	PROA.						
OS	Bacillus sp. 9860.						
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.						
OX	NCBI_TaxID=133778;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=9860;						
RX	MEDLINE=20368675; PubMed=11118284;						
RA	Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,						
RA	Horikoshi K.;						
RT	"Novel oxidatively stable subtilisin-like serine proteases from						
RT	alkaliphilic Bacillus spp.: enzymatic properties, sequences, and						
RT	evolutionary relationships.";						
RL	Biochem. Biophys. Res. Commun. 279:313-319(2000).						
DR	EMBL; AB046403; BAB21266.2; -.						
DR	HSSP; P00782; ISUP.						
DR	InterPro; IPR000209; Peptidase_S8.						
DR	Pfam; PF00082; Peptidase_S8; 1.						
DR	PRINTS; PR00723; SUBTILISIN.						
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.						
DR	PROSITE; PS00138; SUBTILASE_SER; 1.						
SQ	SEQUENCE 639 AA; 68185 MW; 316AF6FFDBE4FF54 CRC64;						
Query Match							
Best Local Similarity 97.2%; Score 2183; DB 2; Length 639;							
Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;							
Qy	1	NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	60	Qy	1	NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	60
Db	206	NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	265	Db	61	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGGLGGLPSNLQTLFSQAYS	120
Qy	61	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGGLGGLPSNLQTLFSQAYS	120	Qy	61	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGGLGGLPSNLQTLFSQAYS	120
Db	266	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGGLGGLPSNLQTLFSQAYS	325	Db	121	AGARIHNTNSWGAAYNGVATTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI	180
Qy	121	AGARIHNTNSWGAAYNGVATTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI	180	Qy	121	AGARIHNTNSWGAAYNGVATTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI	180
Db	326	AGARIHNTNSWGAAYNGVATTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI	385	Db	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSSLPDSSF	240
Qy	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSSLPDSSF	240	Qy	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSSLPDSSF	240
RESULT 3							
Q9AQR0							
ID	Q9AQR0	PRELIMINARY;	PRT;	434	AA.		
AC	Q9AQR0;						
DT	01-JUN-2001 (TReMBLrel. 17, Created)						
DT	01-OCT-2002 (TReMBLrel. 22, Last sequence update)						
DT	01-OCT-2002 (TReMBLrel. 22, Last annotation update)						
DE	Protease (Fragment).						
GN	PROE.						
OS	Bacillus sp. NVL.						
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.						
OX	NCBI_TaxID=133781;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=NVL.						
RX	MEDLINE=20568675; PubMed=11118284;						
RA	Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,						
RA	Horikoshi K.;						
RT	"Novel oxidatively stable subtilisin-like serine proteases from						
RT	alkaliphilic Bacillus spp.: enzymatic properties, sequences, and						
RT	evolutionary relationships.";						
RL	Biochem. Biophys. Res. Commun. 279:313-319(2000).						
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.						
DR	EMBL; AB046406; BAB21269.1; -.						
DR	HSSP; P00782; ISUP.						
DR	InterPro; IPR000209; Peptidase_S8.						
DR	Pfam; PF00082; Peptidase_S8; 2.						
DR	PRINTS; PR00723; SUBTILISIN.						
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.						
DR	PROSITE; PS00138; SUBTILASE_SER; 1.						
KW	Hydrolase; Protease; Serine protease.						
FT	NON_TER 1						
FT	NON_TER 434						
SQ	SEQUENCE 434 AA; 45294 MW; 83517EDDB74125D2 CRC64;						
Query Match							
Best Local Similarity 95.4%; Score 2143; DB 2; Length 434;							
Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;							
Qy	1	NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	60	Qy	1	NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	60
Db	1	NDVARGIVKADVAGSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITAYALGRTN	60	Db	61	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGGLGGLPSNLQTLFSQAYS	120
Qy	61	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGGLGGLPSNLQTLFSQAYS	120	Qy	61	NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSSGGGLGGLPSNLQTLFSQAYS	120
Db	121	AGARIHNTNSWGAAYNGVATTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI	180	Db	121	AGARIHNTNSWGAAYNGVATTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI	180
Qy	121	AGARIHNTNSWGAAYNGVATTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI	180	Qy	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSSLPDSSF	240
Db	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSSLPDSSF	240	Db	181	TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAQGTIFILSARSSLPDSSF	240

300 PNGDOGWGRVTLNKSINVAAYVNEATATATGKATYSFOAQAGKPLKISLVWTDAPGSTTA 359
361 SVTLVNDLDELITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSGTYTIEVQAYN 420
360 SYTLVNDLDELITAPNGKQYVGNDFSYPDNDNDGRNNVNFVINAPOSGTYTIEVQAYN 419
421 VPGPQTFSLAIVN 434
420 VPSGPQRFSLAIVH 433

RESULT 7
Q8T9W1 ID Q8T9W1 PRELIMINARY; PRT; 1825 AA.
AC Q8T9W1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Serine protease/ABC transporter TagD.
GN TAGD.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ax4;
RA Anjard C., Loomis W.F.;
RT "Evolution of the ABC transporters of Dictyostelium.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF466309; AAL/4253.1; -;
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001140; ABC_TM_transporter.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR002029; Peptidase_S8.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR ProDom; PD00006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR ATP-binding; Protease; Transport.
KW ATP-binding; Protease; Transport.
SQ SEQUENCE 1825 AA; 22641 MW; E28160BC78613A3B CRC64;

Query Match 22.8%; Score 511.5; DB 5; Length 1825;
Best Local Similarity 27.9%; Pred. No. 2.2e-21;
Matches 164; Conservative 76; Mismatches 155; Indels 193; Gaps 22;

QY 19 LYQGGQIVAVADTGLDGTGR---NDS-----SMHEAFRGKITALYALGRTNANDTNGH 68
DB 327 LRKGQILSTADTGLDGSCHGFFSDSNPNIPYNSVNLNHRKVVTYIGSL--HDNEDYVDGH 384
QY 69 GTHVAGSVLNG-----STNKGMAPQANLVFQSI-MDSGGGLGGLPSNLQTLFSQAYS 120
DB 385 GTHVCGSAAGAPEDSSIAISSFSGSLATDAKIAFPDLASDPNSNEPVPEDYSQLYQPLYN 444
QY 121 AGARIHTNSWGA-----AVNCAYTDSRNVDDYVKN-DMTILFAGNENGPNGGTISAPGT 175
DB 445 AGARVHGDGWSGLSIQCYLGYSDDAGSIDDFLTHPDIILIRAAAGNEQYSSLLS-QAT 503
QY 176 AKNAITVGATENLRPSF-----
DB 504 AKNVIIVGAQTTHESYTTDALEYSNFETVAKSTLNSLCQSFDDKCYTVTTAQCCTEYST 563
QY 193 -----GSYAD-----NINIVAOFSSRGPTKDGRIKPDVMAPGTFLISA 230
DB 564 VKGLSGCCTSYIKNSYASIFSSQPELYNENNICSFSKGPETHDGLKPKDIPVAPGQVITSA 623
QY 231 RSSIA-----PDSSEWANHDSKYVMGTSMTATPIVAGNVAQLREH----- 271

301 PNGNOGWGRVTLNKSINVAAYVNESSSLSTOKATYSFATATAGKPLKISLVWSDAPASTTA 360
300 PSQDOGWGRVTLNKSINVAAYVNEATATGKATYSFQTAQAGKPLKISLVWTDAPGSTTA 359
361 SVTLVNDLDELITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSGTYTIEVQAYN 420
360 SYTLVNDLDELITAPNGKQYVGNDFSYPDNDNDGRNNVNFVINAPOSGTYTIEVQAYN 419
421 VPGPQTFSLAIVN 434
420 VPSGPQRFSLAIVH 433

RESULT 6
Q9AQR2 ID Q9AQR2 PRELIMINARY; PRT; 433 AA.
AC Q9AQR2;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Protease (Fragment).
GN PRO.
OS Bacillus sp. Y.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y;
RX MEDLINE=20568675; PubMed=11118284;
RA Sasaki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships.";
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046404; BAB21267.1; -;
DR HSP; O45670; IDBI.
DR InterPro; IPR002029; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 433
SQ SEQUENCE 433 AA; 45587 MW; B81291A803C775AE CRC64;

Query Match 88.5%; Score 1987.5; DB 2; Length 433;
Best Local Similarity 87.3%; Pred. No. 2.1e-108;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTCGRNDSMSHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTCGRNDSMSHEAFRGKITALYALGRTN 60
QY 61 NANTDNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NASDPNGHGHVAGSVLNG-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARTHTNSWGAAGVNCAYTTDSRNVDDYVKNRDMTILFAGNENGPNGGTISAPGTAKNAI 180
DB 120 AGARLHTNSWGAAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNENGPNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSADNINHVAQFSRSGPTKDGRIKPDVMAPGTFLISARSSLPDSSF 240
DB 180 TVGATENVRPSFGSIADNPNHIAQFSRSGATDRGRIKPDVTAFTFLISARSSLPDSSF 239
QY 241 WANHDSKYVMGTSMTATPIVAGNVAQLRHFVKNRGTIPKPSLLKAALTAGAADIGLGY 300
DB 240 WANYSKYVMGTSMTATPIVAGNVAQLRHFVKNRGTIPKPSLLKAALTAGAADIGLGY 299
QY 301 PNGNOGWGRVTLNKSINVAAYVNESSSLSTOKATYSFATATAGKPLKISLVWSDAPASTTA 360

```

Db 624 RSNQANTDQCQDGLPNTNALLSE-----SGTSMATPLATAATILLRQYLVDGYPT 676
QY 272 --FVKNRGITPKPSLLKAAIACADIGLGP-----NGNQGWGRV 310
Db 677 GSIVESNKLQPTGSLKALMINNAOLLNGTFPLSSNTNPNSNAVEDTFAGANFVQGWGSL 736
QY 311 TLDKSLNVAAYNESS-----SLTSOKATYSFT-----338
Db 737 RMSEWL---YVESSGVKPKPSRWVGIGELGDKDKKASNWKEYSLSTGQNVSYCFYKPPSS 793
QY 339 --ATAGKP-LKLSLWSDAPASTTASVTLVNDLDLVIAPNGTQ-----YVGNDFTSYND 391
Db 794 GNSGGIPRIVATLVTWTPPSYSGAKLNLVNLDTMT---NTESEFIFYNSGGSSYNG 850
QY 392 N-----WDGRNNVNFV---INAPQSGTGTVEQAYNPVGPOTFS 429
Db 851 TKGTTLPQDSINNVEGIYPIPTNKSEISFRFIAGTNIPGPONFS 898

RESULT 8
Q9GTN7
ID Q9GTN7 PRELIMINARY; PRT; 1702 AA.
AC Q9GTN7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE TAGA.
GN TAGA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
SEQUENCE FROM N.A.
RA "TagA, a putative serine protease/ABC transporter of Dictyostelium
RT that is expressed at the onset of development and is required for the
RT differentiation of a subpopulation of prespore cells.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AF263455; AAG11416.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR001140; ABC_TM_transporter.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00664; ABC_membrane; 1.
DR Pfam: PF00005; ABC_tran; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 1702 AA; 187103 MW; 4A67716303CB7131 CRC64;

Query Match 19.9%; Score 447; DB 5; Length 1702;
Best Local Similarity 25.8%; Pred. No. 1.2e-17;
Matches 140; Conservative 99; Mismatches 164; Indels 140; Gaps 22;

QY 18 GLYGQGIYAVADTGLDTCR---NDSSMHEAFRGKITALYALGRTNMANTNGHGTAVAG 74
Db 300 GLKGDEIVGCADTGIDINHCFFYDTPNIGSTHKKIIS-YSSNGDQIDEGHGTIVG 358
QY 75 SVLNGSTN-----KGMAPQANLVFQSI-MDSGGGLGLPSNLQTLFSAQYAGARIHT 127
Db 359 TIIGSTTVDPSPYSEFSGGAPNSKVAFVDLQVSGNGL-SIQSNLTAIYSTYDONAKVHC 417
QY 128 NSWGAHVNGAYTTDSRNVDYVRKN-DMTILFAAGNEGPNGG-----TISAPGTAKNAIT 181
Db 418 DAWNSNIGPFYTGVTMEMIDRFQWDHDFLWRSAGN-NVNFEGNSIYTLQSSTSKSLV 476
QY 182 VCATENLRPFGSYADNINH-----201
Db 477 VGSSN--QPS-STYLLSSIDDWDFIYNSIRTSVCTQGQSIYGITCSDVPTQTTSVDIQT 533

```

```

QY 202 -----VAQFSRRGPTKDGRIKPDVWAPGTFILS 229
Db 534 QCCSNPILAKICCCSTEIQOQYQTNSTVSEFIPSLFSGVGPTSDGRKLPDLIAPGSPILS 593
QY 230 ARSSLADPSSFWANH-----DSKAYMGGTSMATPIVAGNVAQLRE-----HFVKN 275
Db 594 SR-SLGPSSSTI--NHCSPITSGIATSIAMEGSSQAAAVATSAAVLVRYQYRGYFING 650
QY 276 R-----GITPKPSLLKAAIAGAA---DIGLGPNGNOGWGRVTLTDKSLNVAAYNESS--- 325
Db 651 KVNSSVGFQPSASLVKATLINTASINVDSTLEY---SOGFGNIQLSKLITTTNAOTTSLD 707
QY 326 -----SLTSOKATYSFTATAGKPLKISLWSDAPASTTASVTLVNDLDLVIITA- 374
Db 708 IPSSIEKADPIINTGETNSYCFSLDSKADIDITLVWTDPAQSPPLSTFTLVNLDLALLAF 767
QY 375 PNG--TOYVGNDFTSYNDN---WDGRNNVNFVINAPOSQGTIYIEVQAYNPVGPOTFS 429
Db 768 VDGELSIYSGNSEITFKNTSOVIFDQLNNVEIRIKDAPIGSYDVKIFGTNIVIPNQSYS 827
QY 430 LAI 432
Db 828 VVI 830

RESULT 9
Q8U0C9
ID Q8U0C9 PRELIMINARY; PRT; 654 AA.
AC Q8U0C9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Alkaline serine protease.
GN Pfl670.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE010265; AAL81794.1; -.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 654 AA; 70230 MW; 1CB145A5F505DB34 CRC64;

Query Match 18.4%; Score 414; DB 17; Length 654;
Best Local Similarity 29.8%; Pred. No. 2.8e-16;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;

QY 18 GLYGQGIYAVADTGLDTCRNDSSMHEAFRGKITALYALGRTNMANTNGHGTAVAG 71
Db 154 GYDGGSTIIGIITDGTID-----ASHPDLQGV-----IGWVDFVNGRSPYDDHGHGTH 202
QY 72 VAGSVLNGSTN-----KGMAPQANLVFQSI-MDSGGGLGLPSNLQTLFSAQYAGARI 125
Db 203 VASTAAGTGAASNGYKGMPCAKLAGIKVLGADGSGSTIITIKGVWAVDNKDKYGIKV 252
QY 126 HTNSWA-----AVNGAYTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAP 173
Db 263 INLSGSSQSSDGTDLASQAVNAWDA-----GLVWVAAGNSGPNKYITGSP 310
QY 174 GTAKNAITVGTATENLRPFGSYADNINHVAQFSRRGPTKDGRIKPDVWAPGTFILSARS 233
Db 311 AAASKVITVGA-----VDKYDVTTSFSSRGPTADGRKLPVAVAPGNWIIAARAS 359

```

QY	279	TPKPSLLKAALIAAGAADICGLGYPNGOGWRVTLDKSLNVA-----YVNES 322
DB	406	TPNDA--KNLIMSTAKSWGPPSKNDYGAGRDLGYEAIRVAGNFRGNNDIVPNHYI--S 461
QY	325	SSLSTSOKA-TYSFTAT-AGKPLKISLV---WSDAPASTTASVTLVNDLDTITAPNTQT 379
DB	462	GYPGSRYSDTWFNATNTSPYAITLIIPDWANYP-----DFDIYLDPSGTL 511
QY	380	YVGNDTSPYNWDNNVNFVFINAPQSGTYTIEVOAY 419
DB	512	IKSS-----TGTRQGTIIILPSTGTYYVKVYS 541
RESULT 11		
Q9FBZ4	ID	Q9FBZ4 PRELIMINARY; PRT; 1239 AA.
AC	Q9FBZ4;	
DT	01-MAR-2001	(TREMBLrel. 16, Created)
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)
DE		Putative secreted peptidase.
GN	SCO7188 OR SC8A11.16C.	
OS		Streptomyces coelicolor.
OC	Bacteria; Actinobacteria;	Actinobacteridae; Actinomycetales;
OC	Streptomycineae;	Streptomycetaceae; Streptomyces.
OX	NCBI_TaxID=1902;	
[1]		
RN	SEQUENCE FROM N.A.	
RC	STRAIN=A3(2);	
RA	Saunders D.C., Harris D.;	
RL	Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=A3(2);	
RA	Cerdeno A.M., Parkhill J., Barrell B.G., Denapaitte D., Eichner A., Cullum J.,	
RL	Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=A3(2);	
FX	MEDLINE=97000351; PubMed=8843436;	
RA	Redenbach M., Kieser H.M., Denapaitte D.,	
RA	Kinashi H., Hopwood D.A.;	
RT	"A set of ordered cosmids and a detailed genetic and physical map for	
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome."	
RL	Mol. Microbiol. 21:77-96(1996).	
RN	[4]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=A3(2) / M145;	
RX	MEDLINE=21996410; PubMed=12000953;	
RA	Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,	
RA	Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,	
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,	
RA	Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,	
RA	Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,	
RA	Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,	
RA	Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,	
RA	Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,	
RA	Hopwood D.A.;	
RT	"Complete genome sequence of the model actinomycete Streptomyces	
RT	coelicolor A3(2)."	
RL	Nature 417:141-147(2002).	
DR	EMBL; AL939130; CAC01588.1; .	
DR	HSSP; Q99405; IMPT	
DR	InterPro; IPR003137; PA.	
DR	InterPro; IPR00209; Peptidase_S8.	
DR	Pfam; PF02225; PA; 1.	
DR	Pfam; PF00082; Peptidase_S8; 1.	
DR	PRINTS; PR00723; SUBTILISIN.	
DR	PROSITE; PS00840; PA; 1.	
DR	PROSITE; PS00136; SUBTILASE ASP; 1.	
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.	
DR	PROSITE; PS00138; SUBTILASE_SER; 1.	
RESULT 10		
Q8RBJ2	ID	Q8RBJ2 PRELIMINARY; PRT; 561 AA.
AC	Q8RBJ2;	
DT	01-JUN-2002	(TREMBLrel. 21, Created)
DT	01-JUN-2002	(TREMBLrel. 21, Last sequence update)
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)
DE		Subtilisin-like serine proteases.
APR2	OR TTE0824.	
OS	Thermoanaerobacter tengcongensis.	
OC	Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;	
OC	Thermoanaerobacteriaceae; Thermoanaerobacterales.	
OX	NCBI_TaxID=119072;	
[1]		
RN	SEQUENCE FROM N.A.	
RC	STRAIN=MB4 / JCM 11007;	
RC	MEDLINE=21992816; PubMed=11997336;	
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,	
RA	Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,	
RA	Tan H., Chen R., Wang J., Yu J., Yang H.;	
RA	"A complete sequence of T. tengcongensis genome."	
RL	Genome Res 12:689-700(2002).	
DR	EMBL; AE013049; AM24081.1;	
DR	InterPro; IPR002048; EF-hand.	
DR	InterPro; IPR000209; Peptidase_S8.	
DR	Pfam; PF00082; Peptidase_S8; 1.	
DR	PROSITE; PS00118; EF_HAND; 1.	
DR	PROSITE; PS00136; SUBTILASE ASP; 1.	
DR	PROSITE; PS00137; SUBTILASE_HIS; 1.	
DR	PROSITE; PS00138; SUBTILASE_SER; 1.	
KW	Protease; Complete proteome.	
SEQ	SEQUENCE 561 AA; 59968 MW; BA9C5C52F7083A18 CRC64;	
Query Match	18.1%; Score 406.5; DB 16; Length 561;	
Best Local Similarity	31.3%; Pred. No. 6.3e-16;	
Matches 144; Conservative	62; Mismatches 135; Indels 119; Gaps 22;	
QY	6	GIYKADVAOSSYGLYGQGOTVAVADTGLDGRNDSSMHEAFRGKITAYALGRTNNA-- 63
DB	155	GITK----ARSDFGVTKNIITAIITDIDGNHVDLS-----GGKI-----IGWKDFINN 201
QY	64	----DTNGCHTHVAGSVLNGNSTN---KGMAPOANLVFQSTMDSGG-----LGGLPSN 110
DB	202	TFPYDDMGHTHWASIAAGTGAGNSFYKGVAPDALLVIKVIDANGSSMTVTAGIDWA 261
QY	111	LQ-----TLFSQAYSAGRIHTNSWGAAVNGYTDSNRNVDDYVRKKNDMTILF 158
DB	262	VQNKDVIYGIKVNLSLGTSSTDG-----TDSTSLAVN-----RAVD-----SGIVVV 305
QY	159	AAGNEGPNGGTISAPGTAKNATVCATENLRPSFGSYADNHNHVAQSFSSRGPTKDGRIKP 218
DB	306	AAGNSGPACTYIGPSGAERKAITVAAMADV-----GELGFNL-----ASFSSRGPTADGRKIP 358
QY	219	DVMAPGFFILSARSLAPDGSFWANHDPSKYAYMGCTSMATPIVAGCNVAQLREHFVKNRGI 278
DB	359	DIAPGNYNTAAK-----ANSVNGYVYTSCTSMATPFVAGTVAGTMLN-----ANPNL 405

Query Match	17.0%	Score 381;	DB 16;	Length 430;
Best Local Similarity	34.8%;	Pred. No. 1.4e-14;		
Matches 11;	Conservative 53;	Mismatches 126;	Indels 40;	Gaps 16;
QY	2	DVARGIVKADVAQSSGLYQGOIVAVADTGLTGRNDSMHEAFGRKIT--ALYALGRT	59	
Db	121	DTASSINADVLKES-GLTQGGSTIAVIDTGIHP-----HEDLEGRITGFADFVKGT	172	

```

DR Pfam: PF02225; PA: 1.
DR MEROPS: S08.069; -.
DR InterPro: IPR002860; GH_BNR.
DR InterPro: IPR00209; Peptidase_s8.
DR Pfam: PF02012; BNR: 2.
DR PROSITE: PS00840; PA: 1.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Complete proteome.
SQ SEQUENCE 1253 AA; 130971 MW; AA69B417EEFEDB89 CRC64;

Query Match          16.7%; Score 376; DB 16; Length 1253;
Best Local Similarity 27.6%; Pred. No. 1.1e-13;
Matches 142; Conservative 58; Mismatches 145; Indels 170; Gaps 22;

QY 16 SYGLYGQGIIVAVADTGLDTRNDSSMHEAFKITALVALGRTNNDNTHGHVAGS 75
DB 233 SGGNTGEGVGAVLDTGVDA-----HPDFAGRIATAATSFVDPQDVTDRNGHGHVAST 286
QY 76 VLNGSTN-----KGMAPQANLYFQSIIMDSGG-----CLGG 106
DB 287 VAGTGAASGGVEKGVAPGASLHIGKVLNDSGQDSWVLGMEWAVRDQHAKIVMSLGD 346
QY 107 LPSNLQTLFSQAYSAGARIHTNSWGAAVNGAYTTDSRNVDYVYRKNDMTILFAAGNEGN 166
DB 347 SPTDGTDLSEAV-----NLSAETGA-----LFVVAAGNSGPE 380
QY 167 GGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAQFSRGP-TKDGRIKPDVMAFGT 225
DB 381 AYTGTTPAAADAALTVGAVNG-PGKG-----VDQLADFSSRGPVRVDNAVKPDLTAPGV 433
QY 226 FILSARSLAPSSFWANHDSKYAYMGTSMATPIVAGNVAQLREHVEVKNRGIPTKPSLL 285
DB 434 GYLAARSYAPG-----EGAYQSUSGTSMATPHVAGAAALAAEHPDWTG-----QRL 482
QY 286 KAALIAAGADIIGLYPENGQ-----GWRVTLDKSLNVAIYVNESSLSSTQ---KATY 335
DB 483 KEALVCTTA-----GTQRESFPDAGSGRV-----DVAAAVRSTLLASGDFAQAAY 528
QY 336 -----SFTATAGKPLKISLVSDA-----PASTTASVTLVND 367
DB 529 PYTPGQTVRRDVTYITNSGPAPVALDLALSPAELPEGLFTTSEAQVTPAHGTASVGVITH 588
QY 368 LD-----LVITAPNGTOVVGNDFTSPYNDNWDGRNNVNFINA-----POSG 410
DB 589 LDAAEDNGAYATRLVAGSADGAVLA---RTPVGVNKEGRR--ATLALTAKDHHDKPLSG 642
QY 411 T-----YTIEVQA-YNVPVGPQTFSL 430
DB 643 TVILKDVERNTAPKVYSVDASGRLLDLRLSPSTYSV 677

RESULT 14
P95684 ID ID PRELIMINARY; PRT; 1102 AA.
AC AC P95684;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Streptomyces albogriseolus.
DE Subtilisin-like protease.
OS Streptomyces albogriseolus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-3253;
RX MEDLINE=97144528; PubMed=8990295;
RA Suzuki M., Taguchi S., Yamada S., Kojima S., Miura K., Momose H.;
RT "A novel member of the subtilisin-like protease family from
Streptomyces albogriseolus."
RL J. Bacteriol. 179:430-438(1997).
CC -/- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DB EMBL; D83672; BAA12040.1; -.

```

```

DR HSP: P00782; 2SBT.
DR MEROPS: S08.069; -.
DR InterPro: IPR002860; GH_BNR.
DR InterPro: IPR00209; Peptidase_s8.
DR Pfam: PF02012; BNR: 2.
DR PROSITE: PS00840; PA: 1.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 1102 AA; 114128 MW; F9E4AD2590FE59E CRC64;

Query Match          16.3%; Score 366; DB 2; Length 1102;
Best Local Similarity 31.6%; Pred. No. 3.6e-13;
Matches 148; Conservative 55; Mismatches 175; Indels 90; Gaps 20;

QY 3 VAR-----GIVKADVAQS-----SYGLYGQGIIVAVADTGLDTRNDSSMHEAFPG 48
DB 183 VARVWLDGVRKASLDTSVGQIGTPKAWEGYDGKGVKIAVLDTGVD-----ATHPDLKG 236
QY 49 KITALVALGRTNNDNTHGHVAGSVLNGS-----TNKGMAPQANLYFQSIIMDSGGGL 104
DB 237 QVTASKNFTSAPTTGVDVVGHTHVASTAAGTGAOSKGTGKGVAPGAKILNGKVLDDAG-- 294
QY 105 GGLPSNLQTLFSQAYSAGARIHTNSWGAAVNGAYTTDSRNVDYVYRK--NDMTILF--AA 160
DB 295 FGDDSGILAGMEWAAAGADIVNMSLG---GMDTPETDPLEAAVDKLSAEKGIILFAIAA 350
QY 161 GNEGPGNGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAQFSRGP-TKDGRIKPD 219
DB 351 GNEGPQ--STGSPGSDASLTGVA-----VDDKDLADFSSTGPRLCDGAVKPD 397
QY 220 VMAPGTFILSARSLAPSSFWANHDSKYAYMGTSMATPIVAGNVAQLREHVEVKNRGI 279
DB 398 LTPAGVDITAAAKGNDIAKEVGEKPGAGYMTISGTSMATPHVAGAAALLKQOHP----- 452
QY 280 KPESLLKAALIAAGADIGLG-YPNQNGWGRVTLDKSLNVAIYVNESSLS----- 328
DB 453 WKYAEKLGALTASTKD---GRYTPFEQSGRQVQVDKAITQTVAIEFVSUSFGVQOQPHAD 509
QY 329 ---TSOKATYSFTATAGKPLKISLVSDA-----APAS---TTASVTLVNDLVLITAP-NG 377
DB 510 DKPVTKRLTNRNLCTEDVTLLKLTSTATGPKAAAGPFTLGASLT-----TVPANG 561
QY 378 TQYVGNDFTSYNDNWDGRNNVNFINAPOS---GTYTIEVQAYNV 421
DB 562 TASVDVTADTFLGCAVDGTYSAYVAVATGAGQSVRTAAAVEREVESYV 609

RESULT 15
Q9RL54 ID ID PRELIMINARY; PRT; 1245 AA.
AC AC Q9RL54;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable secreted peptidase.
GN SC00432 OR SCF51A.10.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RC Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RL

```

Search completed: July 25, 2003, 19:01:25
Job time : 41.0404 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:59:51 ; Search time 21.0242 Seconds
(without alignments)
2451.541 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EYQAYNPVPGQTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pcp.*
- 2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pcp.*
- 3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pcp.*
- 4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pcp.*
- 5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pcp.*
- 6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pcp.*
- 7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pcp.*
- 8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pcp.*
- 9: /cgn2_6/ptodata/2/pubaa/US09A_PUBCOMB.pcp.*
- 10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pcp.*
- 11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pcp.*
- 12: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pcp.*
- 13: /cgn2_6/ptodata/2/pubaa/US10A_PUBCOMB.pcp.*
- 14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pcp.*
- 15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pcp.*
- 16: /cgn2_6/ptodata/2/pubaa/US10_NEW_PUB.pcp.*
- 17: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pcp.*
- 18: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2247	100.0	434	11	US-09-985-689A-1
2	2191	97.5	434	11	US-09-985-689A-2
3	2143	95.4	434	11	US-09-985-689A-6
4	2125.5	94.6	433	11	US-09-985-689A-7
5	1998.5	88.9	433	11	US-09-985-689A-5
6	1994.5	88.8	433	11	US-09-985-689A-3
7	1987.5	88.5	433	11	US-09-985-689A-4
8	452.5	20.1	659	14	US-10-090-624-12
9	414	18.4	412	14	US-10-090-624-1
10	414	18.4	522	14	US-10-090-624-1
11	414	18.4	654	14	US-10-090-624-16
12	366	16.3	1079	15	US-10-112-488-39
13	363.5	16.2	1208	15	US-10-156-761-13251
14	346.5	15.4	1139	15	US-10-156-761-10856
15	304.5	13.6	1398	14	US-10-090-624-6

16	283	12.6	580	11	US-09-927-827-55
17	280.5	12.5	595	11	US-09-927-827-59
18	276	12.3	418	10	US-09-966-921A-2
19	270	12.0	1101	15	US-10-156-761-12934
20	251	11.2	280	15	US-10-209-812-2
21	251	11.2	397	11	US-08-779-334A-5
22	247	11.0	271	11	US-09-813-408-2
23	246.5	11.0	379	11	US-09-813-408-6
24	242.5	10.8	627	11	US-09-927-827-60
25	240	10.7	269	8	US-08-322-678-10
26	240	10.7	269	9	US-09-837-235-16
27	240	10.7	269	9	US-09-060-854B-6
28	240	10.7	269	10	US-09-975-139-1
29	240	10.7	269	10	US-09-976-414-8
30	240	10.7	269	14	US-10-075-907-1
31	240	10.7	269	14	US-10-075-895-1
32	240	10.7	269	15	US-10-033-325-6
33	240	10.7	269	15	US-10-209-812-3
34	240	10.7	269	15	US-10-104-693-4
35	240	10.7	269	15	US-10-228-572-6
36	240	10.7	377	11	US-09-813-408-1
37	240	10.7	380	11	US-09-824-893A-261
38	238.5	10.6	275	15	US-10-104-693-3
39	238.5	10.6	279	11	US-09-813-408-21
40	237.5	10.6	382	15	US-10-033-325-2
41	237.5	10.6	382	15	US-10-228-572-2
42	236	10.5	266	9	US-09-837-235-18
43	236	10.5	269	11	US-09-813-408-4
44	236	10.5	380	11	US-09-813-408-19
45	235.5	10.5	275	8	US-08-322-678-7

ALIGNMENTS

RESULT 1
US-09-985-689A-1
; Sequence 1, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-1

Query Match	100.0%;	Score 2247;	DB 11;	Length 434;
Best Local Similarity	100.0%;	Pred. No. 3e-191;		
Matches 434;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTCGRNDSMHAEFRGKITALYALGRN	60	
Db	1	NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTCGRNDSMHAEFRGKITALYALGRN	60	
QY	61	NANDTNGHGTAVAGSVLNGSTNKGMAFOANLVFQSIMDSGGGLGSLPSNLQTLFQAYS	120	

Db 61 NANTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 300
Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGPKPLKISLWSDAPASTTA 360
Db 301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGPKPLKISLWSDAPASTTA 360
QY 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQGTYYIEVOAYN 420
Db 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQGTYYIEVOAYN 420
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 2

US-09-985-689A-2

; Sequence 2, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483050
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.

US-09-985-689A-2

Query Match 97.5%; Score 2191; DB 11; Length 434;
Best Local Similarity 96.5%; Pred. No. 2,8e-186;
Matches 419; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NANTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180

QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 300
Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGPKPLKISLWSDAPASTTA 360
Db 301 PNGNOGWRVTLDKSLNVAYNVNESSLSSTOKATYSFTATAGPKPLKISLWSDAPASTTA 360
QY 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQGTYYIEVOAYN 420
Db 361 SVTLVNDLDLVTAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQGTYYIEVOAYN 420
QY 421 VPVGPQTFSLAIVN 434
Db 421 VPVGPQTFSLAIVN 434

RESULT 3

US-09-985-689A-6

; Sequence 6, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483050
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.

US-09-985-689A-6

Query Match 95.4%; Score 2143; DB 11; Length 434;
Best Local Similarity 93.5%; Pred. No. 5.2e-182;
Matches 406; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NANTNGHGHVAGSVLNGSTNKGMAPQANLVFQSDSGGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 300
Db 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADIGLY 300

[illegible]

```

RESULT 4
US-09-985-689A-7
; Sequence 7, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483050
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-7

```

[illegible]

```

Db      360  SLTLVNDLDELITAPNGTKYVGNDETFAPYDNWDGRNNVENFIAPOSGETTYVEQAYN 419
Qy      421  VPVGFQTFSLAIVN 434
          ||| |||||
Db      420  VPVGFQTFSLAIVH 433

RESULT 5
US-09-985-689A-5
; Sequence 5, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-5

```

Query Match	88.9%;	Score 1998.5;	DB 11;	Length 433;
Best Local Similarity	87.8%;	Pred. No. 3.6e-169;		
Matches 381;	Conservative 28;	Mismatches 24;	Indels 1;	Gaps 1;
Qy	1	NDVARGIVKADVAOSSGLYGOGQIVAVADTGLDTGRNDSMHAEFRGKITALYALGRTN	60	
Db	1	NDVARGIVKADVAONNYGLYGQGVAVADTGLDTGRNDSMHAEFRGKITALYALGRTN	60	
Qy	61	NANDNGHGHVAGSVGLNGSTNKGMAPOANLVFQSIMDSGGJGLGSPNLQTLFSQAYS	120	
Db	61	NANDPNGHGHVAGSVGLN-ALNKGMAPOANLVFQSIMDSGGJGLGSPNLQTLFSQAWN	119	
Qy	121	ACARHTHSWGAAGVNCAYTTTDSRNVDYVFKNDWTLTFAAGNEGPNGGTISAPGTAKNAI	180	
Db	120	ACARHTHSWGAAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNEGPNSGTISAPGTAKNAI	179	
Qy	181	TVGATENLRPSPFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTFFILSARSSLAPDSSE	240	
Db	180	TVGATENLRPSPFGLADNPNHIAQFSSRGATRDGRIKPDVTA PGTFILSARSSLAPDSSE	239	
Qy	241	WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADIGLGY	300	
Db	240	WANYNSKYAYMGGTSMATPIVAGNVAQLREHF IKNRGITPKPSLIKALTAGATDVGLGY	299	
Qy	301	PNGOQGWGRVTLDKSLNVAIVNNESSLSSTOKATYSFTATAGPKLKSILSWSDAPASHTA	360	
Db	300	PSGQGWGRVTLDKSLNVAIVNREATALATGQATYSFQAQAGPKLKSILSWWDAGSTTA	359	
Qy	361	SVTLVNDLDLVITAPNGTQYVGNDDTSPSYDNNDWGRNNVENVFINAPQSGTYTTEVOAYN	420	
Db	360	SYTLVNDLDLVITAPNGOKYVGNDFSYPYDNNDWGRNNVENVFINAPQSGTYTTEVOAYN	419	
Qy	421	VPVGPQPTFSIAVN	434	
Db	420	VPSPQPTFSIAIVH	433	

Tue Jul 29 14:12:42 2003

```

; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-4

```

```

Query Match      88.8%; Score 1987.5; DB 11; Length 433;
Best Local Similarity 87.3%; Pred. No. 3.4e-168;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

```

```

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAQANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NASDPNGHGHVAGSVLGN-ALNKGMAQANLVFOSIMDSGGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVYRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAANGAYTTDSRNVDYVYRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPTFTLSARSSLAPDSSF 240
Db 180 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVMAPTFTLSARSSLAPDSSF 239
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLIKAALIAGAADIGLY 300
Db 240 WANTNSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLIKAALIAGAADIGLY 299
QY 301 PNGNQGWGRVTLDKSLNVAIVYVNESSLSSTOKATYSFTATAGPKPLKISLVWSDAPASTTA 360
Db 300 PNGDQGWGRVTLDKSLNVAIVYVNEATATGOKATYSFOAQAGPKPLKISLVWTDAPGSTTA 359
QY 361 SVTLVNDLVLITAPNGTOYVGNDFTSYNDNDGRNNVNFVFNAPQSGTYTIEVQAYN 420
Db 360 SVTLVNDLVLITAPNGQYVGNDFTSYNDNDGRNNVNFVFNAPQSGTYTIEVQAYN 419
QY 421 VPGVQPTFSLAIVN 434
Db 420 VPSGQRFSLAIVH 433

```

```

RESULT 8
US-10-090-624-12
; Sequence 12, Application US/10090624
; Publication No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06

```

```

; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-3

```

```

Query Match      88.8%; Score 1994.5; DB 11; Length 433;
Best Local Similarity 87.6%; Pred. No. 8e-169;
Matches 380; Conservative 28; Mismatches 25; Indels 1; Gaps 1;

```

```

QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITIYALGRTN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAQANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NASDPNGHGHVAGSVLGN-ALNKGMAQANLVFOSIMDSGGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARIHTNSWGAANGAYTTDSRNVDYVYRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAANGAYTTDSRNVDYVYRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPTFTLSARSSLAPDSSF 240
Db 180 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVMAPTFTLSARSSLAPDSSF 239
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLIKAALIAGAADIGLY 300
Db 240 WANTNSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLIKAALIAGAADIGLY 299
QY 301 PNGNQGWGRVTLDKSLNVAIVYVNESSLSSTOKATYSFTATAGPKPLKISLVWSDAPASTTA 360
Db 300 PNGDQGWGRVTLDKSLNVAIVYVNEATATGOKATYSFOAQAGPKPLKISLVWTDAPGSTTA 359
QY 361 SVTLVNDLVLITAPNGTOYVGNDFTSYNDNDGRNNVNFVFNAPQSGTYTIEVQAYN 420
Db 360 SVTLVNDLVLITAPNGQYVGNDFTSYNDNDGRNNVNFVFNAPQSGTYTIEVQAYN 419
QY 421 VPGVQPTFSLAIVN 434
Db 420 VPSGQRFSLAIVH 433

```

```

RESULT 7
US-09-985-689A-4
; Sequence 4, Application US/09985689A
; Publication No. US2003002235A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI

```


18 GYDGGGIVAVADTGLDTRNDSSMHEAFRGKITALYALGRNTNAN-----DTNGHGTH 71
 22 GYDGGGIVAVADTGLDTRNDSSMHEAFRGKITALYALGRNTNAN-----DTNGHGTH 70
 72 VAGSVLNGSTN-----KGMAPQANLVFQSIM--DSGGGLGLPSNLQTLFQSAQSAGARI 125
 71 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVENAVDNKDKYGIKV 130
 126 HTNSWGA-----AVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNCGTISAP 173
 131 INLSGSSQSDGTDALSOAVNAWDA-----GLVVVAAGNSGPNKYITGSP 178
 174 GTAKNATVGTATENLRFSGYADNINHVAFSSRGPTKDGRIKPDVMPAGTIFLSARSS 233
 179 AAASKVITVGA-----VDKYDITFSRSSGPTADGRUKPEVAVPAGNWIIAARAS 227
 234 LAPDSSFWANHDSKYAYMGSTPMATPIVAGNVAQLREHFVKRGIPTK--PSLLKAALIA 291
 228 ---GTSMGQPIINDYITAAFGTSMATPHVAGIAALLQ-----AHPSWTPDKVKTALIE 277
 292 GA-----ADIGLYPENGNGWRVTLDKSLNVAIVNNESSLSSTOKA-----TYSFT 338
 278 TADIVKPEIADIAIGA-----GRVNAIKAIN--YDNYAKLVFTGYVANKGSOHOFV 328
 339 ATAGKPLKISLWSDAPASTTASVTLVNDLDLITAPNGTQYVGNDFTSFYNDNWDGRNN 398
 329 ISGASFVATLWDMAN-----SDDLILYDPNGNQ--VDYSYTAAY-----G 369
 399 VENFINAPQSGTITIEVOAYNVPGPQTFSLAIVN 434
 370 FEKVGYNPTDGTWIKVVSYS---GSANYQVDVVS 402

RESULT 11

US-10-090-624-16
 ; Sequence 16, Application US/10090624
 ; Publication No. US20020132335A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAKURA, Hikaru
 ; APPLICANT: MORISHITA, Mio
 ; APPLICANT: SHIMOJO, Tomoko
 ; APPLICANT: ASADA, Kiyoaki
 ; APPLICANT: KATO, Ikunoshin
 ; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
 ; FILE REFERENCE: TAKAKURA=6
 ; CURRENT APPLICATION NUMBER: US/10/090,624
 ; CURRENT FILING DATE: 2002-03-06
 ; PRIOR APPLICATION NUMBER: 09/445,472
 ; PRIOR FILING DATE: 1999-12-06
 ; PRIOR APPLICATION NUMBER: 151969/1997
 ; PRIOR FILING DATE: 1997-06-10
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 16
 ; LENGTH: 654
 ; TYPE: PRP
 ; ORGANISM: Pyrococcus furiosus
 US-10-090-624-16

Query Match 18.4%; Score 414; DB 14; Length 654;
 Best Local Similarity 29.8%; Pred. No. 3.7e-28; Indels 114; Gaps 19;
 Matches 136; Conservative 59; Mismatches 147
 18 GYDGGGIVAVADTGLDTRNDSSMHEAFRGKITALYALGRNTNAN-----DTNGHGTH 71
 154 GYDGGGIVAVADTGLDTRNDSSMHEAFRGKITALYALGRNTNAN-----DTNGHGTH 202
 72 VAGSVLNGSTN-----KGMAPQANLVFQSIM--DSGGGLGLPSNLQTLFQSAQSAGARI 125
 203 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVENAVDNKDKYGIKV 262
 126 HTNSWGA-----AVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNCGTISAP 173

263 INLSGSSQSDGTDALSOAVNAWDA-----GLVVVAAGNSGPNKYITGSP 310
 174 GTAKNATVGTATENLRFSGYADNINHVAFSSRGPTKDGRIKPDVMPAGTIFLSARSS 233
 311 AAASKVITVGA-----VDKYDITFSRSSGPTADGRUKPEVAVPAGNWIIAARAS 359
 234 LAPDSSFWANHDSKYAYMGSTPMATPIVAGNVAQLREHFVKRGIPTK--PSLLKAALIA 291
 360 ---GTSMGQPIINDYITAAFGTSMATPHVAGIAALLQ-----AHPSWTPDKVKTALIE 409
 292 GA-----ADIGLYPENGNGWRVTLDKSLNVAIVNNESSLSSTOKA-----TYSFT 338
 410 TADIVKPEIADIAIGA-----GRVNAIKAIN--YDNYAKLVFTGYVANKGSOHOFV 460
 339 ATAGKPLKISLWSDAPASTTASVTLVNDLDLITAPNGTQYVGNDFTSFYNDNWDGRNN 398
 461 ISGASFVATLWDMAN-----SDDLILYDPNGNQ--VDYSYTAAY-----G 501
 399 VENFINAPQSGTITIEVOAYNVPGPQTFSLAIVN 434
 502 FEKVGYNPTDGTWIKVVSYS---GSANYQVDVVS 534

RESULT 12

US-10-112-488-39
 ; Sequence 39, Application US/10112488
 ; Publication No. US20030082746A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KIKUCHI, Yoshihiko
 ; APPLICANT: DATE, Masayoshi
 ; APPLICANT: UMEZAWA, Yukiko
 ; APPLICANT: YOKOYAMA, Keiichi
 ; APPLICANT: NATSUI, Hiroshi
 ; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSLUTAMINASE
 ; FILE REFERENCE: 219286USOCONT
 ; CURRENT APPLICATION NUMBER: US/10/112,488
 ; CURRENT FILING DATE: 2002-04-01
 ; PRIOR APPLICATION NUMBER: PCT/JP00/06780
 ; PRIOR FILING DATE: 2000-09-29
 ; PRIOR APPLICATION NUMBER: JP2000-280098
 ; PRIOR FILING DATE: 2000-06-28
 ; PRIOR APPLICATION NUMBER: JP11-280098
 ; PRIOR FILING DATE: 1999-09-30
 ; NUMBER OF SEQ ID NOS: 70
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 39
 ; LENGTH: 1079
 ; TYPE: PRP
 ; ORGANISM: Streptomyces albobogiseolus
 US-10-112-488-39

Query Match 16.3%; Score 366; DB 15; Length 1079;
 Best Local Similarity 31.6%; Pred. No. 1.4e-23;
 Matches 148; Conservative 55; Mismatches 175; Indels 90; Gaps 20;
 3 VAR---GIVKADVAQS-----SYGLYGOGQIVAVADTGLDTRNDSSMHEAFRG 48
 160 VARWLDGVRKASLDTSVGQIGTPKAWGAGYDVGKVKIAVLDTGVD-----ATHPDLKG 213
 49 KITAXALGRNTNANDTNGHGTTHVAGSVLNGS-----TNKGMAPQANLVFQSIMDSGGGL 104
 214 QVTASKNFTSAPTPTGDDVVGHTHVASIAAGTGAOSKGYKGVAPGAKILNGKVLDDAG-- 271
 105 GGLPSNLQTLFQSAQSAGARIHTNSWGAAGVAVTDSRNVDDYVRK--NDMTILF--AA 160
 272 FGDDSGILAGMEWAAAGADLVNLSG-----GMDTPTDPLEAAVDKLSAEKILFAIAA 327
 161 GNEGPNCGTISAPGTAKNATVGTATENLRFSGYADNINHVAFSSRGPT--TKDGRKPD 219
 328 GNEGPN--SIGSPGSAADSALTVA-----VDDKDKLADFSSTGPRLDGDAVKPD 374
 220 VMAPGTIFLSARSSLAPDSSEFWANHDSKYAYMGSTPMATPIVAGNVAQLREHFVKRGIPT 279

Db 375 LTAPGVDTAASAKGNDIAKEVGEKPGAGYWTISGTSMAHPVAGAAALLKQOHP----- 429
QY 280 KPSSLKKAALACADIGL-YPNGNQGGRVTLDKSLNVAIVNNESSLS----- 328
Db 430 WKYAEKLGALTASTKD---GKVTPEQSGRGVQDKALTQTVAIEPVSLSFGVQOWPHAD 486
QY 329 ---TSQATYSFTAGPKPLKISLVMSD-----APAS---TTASVTLVNDLVLITAP-NG 377
Db 487 DKPVTKLTYNLGTEDVTLKLTSTATGPKGAAPAGFTTLGASTL-----TVPANG 538
QY 378 TQYVGNDFTSYNDWNGRNVNVPINAPQS-----GTYTIEVOAYNV 421
Db 539 TASVDVTADTRLGGAVDGTYSAYVVATGAGQSVRTAAAVEREVESYV 586

RESULT 13

US-10-156-761-13251
: Sequence 13251, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: PRIOR FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 13251
: LENGTH: 1208
: TYPE: PRT
: ORGANISM: Streptomyces avermitilis
US-10-156-761-13251

Query Match 16.2%; Score 363.5; DB 15; Length 1208;
Best Local Similarity 32.0%; Pred. No. 2.8e-23;
Matches 128; Conservative 54; Mismatches 145; Indels 73; Gaps 18;
QY 8 VRADVACSS-----YGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALG 57
Db 186 VEADMAESNAQIGTRAANDAGLGDGVTVAVLDTGVDT-----THPDLAQRVSRKSFI 239
QY 58 RTNNANDTNGCHTHVAGSVLNGCS-----TNKGAPQANLVFQSTM-DSGGGLGLPSNLQ 112
Db 240 DGEVADRNGHGHVHTSTVGGSGAASDGTGERGAPGATLAVGKVLSDQAG-----SESQ 294
QY 113 TLFQAYSAA---GARLHNSWGA--AVNGAYTDD--SRNVDDYVRKNDMTILFAAGNEGP 165
Db 295 IIAEMEWAAARDVRAIVSLSLSTEASDG---TDPMAEAVDTLSEETGALFVVAAGNTGA 351
QY 166 NGGTSAPCTAKNAITVGTATENLRPSFGSYADNINHVAQFSSRGPTK-DGRIRKPDVMAPG 224
Db 352 -PSSIGSPGAADSALTVA---VDSDDRAAYFTSAGPRHGNALPDLAAGP 399
QY 225 TFLSARSLSPDSFWANHDSKYAYMGTSMAPTPVAGNVAOLREHFVKNRGTITPKPSL 284
Db 400 VDIAARSQLAPGTG-----YTSMSGTSMAPTPVAGNVAOLREHFVKNRGTITPKPSL 452
QY 285 LKAA--LIAGAADIGLGPNGOGWG-RVTLDKSLNVAV-----VNSSSLSTSQ 331
Db 453 MSTSEQLDASYQLGAGRVSVDPVAGARVATGSDGLGFHRPHDADRPVTKTIVTSNNS 512
QY 332 KATYSFT-ATAGKPKLISLVMSD-----PASTTASVTLVND 367
Db 513 DTTVELSLAVRGAPAGVATLADTALTPAHGTAATVTTGDD 552

RESULT 14

US-10-156-761-10856
: Sequence 10856, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: PRIOR FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 10856
: LENGTH: 1139
: TYPE: PRT
: ORGANISM: Streptomyces avermitilis
US-10-156-761-10856

Query Match 15.4%; Score 346.5; DB 15; Length 1139;
Best Local Similarity 29.0%; Pred. No. 8.2e-22;
Matches 143; Conservative 54; Mismatches 179; Indels 117; Gaps 20;
QY 18 GLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRNTNNDTNGHGTAVAGSVL 77
Db 249 GYDGKCVKIAVLDTGVD-----ATHPDLKDQVAESKNFSNAAADAHFGHGTIVASTAA 302
QY 78 GNGSTN----KGMAPQANLVFQSIMDSGGGLGLPSNLQTLFQSOAYSAGARIHTNSWGAA 133
Db 303 GTGAKSNGKYKGVAPGATILNGKVLDDTG--SGDDSGILAGMEWAAEQGADVNLSLG-- 358
QY 134 VNGAYTTDSRNVDDYVRK--NDMTILF--AAGNEGPN--GTISAPCTAKNAITVGTATENL 188
Db 359 --GGDIPEIDPLEAEVKNLSEKGIILFAAAGNEGEFGEOTIGSPGAADALIVGA---- 412
QY 189 RPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLSPDSFWANHDSK- 247
Db 413 -----VNDSDKLSASFSSRGPLDGAIRKPDVTPAGVDITAA---AAPGSVIDQEVGQKP 462
QY 248 --YAYMGTSMAPTPVAGNVAOLREHFVKNRGTITPKPSL--LKAALIAGAADIGLGPNG 303
Db 463 DGYLTISGTSMAPTPVAGAAAILKQOH-----PNWSFAELKGALTGSAR--GGKYTF 513
QY 304 NOGWRVTLDKSLNVAIVNNESSLS-----TSOKATY-----SPTAT 340
Db 514 QQSGRIADVKAIKQSVIANPNSVSGIQOWPHITDDKPYTOOLTLYRNLGTSVTLNLAST 573
QY 341 AKGP-----LKISLWSDAPASTTASVTLVNDL--VITAPNGTQYV 381
Db 574 ATNPKGVAAPSGFFKLGATKVTVPAGKASVDFTVNTKLGTTDGAISAVTATGGQTV 633
QY 382 -----GNDFTSYNDWNGRNVN-----ENVFIPAQSGTITIEV 416
Db 634 RTAAAVQREVESYDVLKHLDR-DGKPAVNYSTDLTGVSGLAADKWFAPYDASGTIVKVR 692
QY 417 QAYNVVPQPTFS 429
Db 693 PKGNFILNASLFA 705

RESULT 15

US-10-090-624-6
: Sequence 6, Application US/10090624

Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 1398
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
; US-10-090-624-6

Query Match 13.6%; Score 304.5; DB 14; Length 1398;
Best Local Similarity 26.6%; Pred. No. 6e-18;
Matches 139; Conservative 57; Mismatches 178; Indels 149; Gaps 22;

Qy 21 GGGTIVAVADTGLDGRNDS-----SMHEAFRGKITAIYALGRTNNAN----- 63
Db 301 GNGYDIAYVDTDLDTDFDEVPLGOYNVTYDVAVFSSYYGPLYVLAIEDPNGEYAVFGV 360

Qy 64 DTNGHGHVAGSVLNGSTN-----KGMAPQAN 91
Db 361 DGHGHGHVAGSVAGVAGYDSNNDWDWLSMYSGEWEVESRLYGWDTYTNVTTDTVOGVAPGAQ 420

Qy 92 LVFOSIMPSGGGLGLPNSLQTLFSQAYSAGARIHTNSGAAVNGAYT--TDSRN--VDD 147
Db 421 IMATRVLS--DGRGSMWDIEGM--TYAATHGADVISMSLGG--NAPYLDGTDPSVAUDE 476

Qy 148 YVRNDMTILFAAGNEGNGGTISAPGTAKNATTVGATENLRPSFGSYAD----- 197
Db 477 LTERGVVVFVIAAGNEGPGINIVGSPGATKAITVGAAA--VPINVGYYVQALCYPDY 535

Qy 198 -----NINHAQFSSRGPKDGRKPKDVMAPGTFILSARSRLAPDSSFWANHDSKYA 249
Db 536 FYFFPAYTNV-RIAFFSSRGPRIDGEIKPNVAVPGYGIYSSLPWIGGADF----- 585

Qy 250 YNGGTSMATPTVAGNVVAQLREHFVKNRGITPKPSLLKAALIAGAADI-----CLGYPNG 303
Db 586 -MSGTSMATPHVSGVVALLISG-AKAEGIYINPDITKKVLESGATWLEGDPYTGOKYTEL 643

Qy 304 NQGWGRVTLDKSLNAYVNESSLSSTSQKATYSFTATAGKPLKISLVWSOAPASTTA--- 360
Db 644 DQCHGLVNVTKSWEI-----LKAINGTITLPVDHWADKSYSDFAEYL 685

Qy 361 SVTLVNDLVLITAPN-----GTYQYGN-----DFTSPYNDWN-----DG-----RNNVENVF 403
Db 686 GVDVIRGLYARNIPDIVEMHIKYGVDTEYRTFEIYATEPWIKPFVSGSVILENTEFVL 745

Qy 404 -----INAPQSGTY-----TIEVOAYNVVPGPOTFS 429
Db 746 RVKYDVEGLEPGLVGRRIIDPTTPVIEDEILNTIPIKEFT 788

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 28, 2003, 03:38:50 ; Search time 2901.34 seconds
(without alignments)
3635.607 Million cell updates/sec

Title: US-09-985-689A-1
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQYNVPVGPQTFLAIVN 434

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frames_p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09985689/runat_25072003_143032_845/app_query.fasta_1.1166
-DB-EST-QFMT-fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cd1 -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @Cn1_1_5436@runat_25072003_143032_845 -NGPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_esthum.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	246	10.9	1605	13	BQ622771	BQ622771 CC_Contig
2	241	10.7	640	12	BJ395336	BJ395336 BJ395336
3	225.5	10.0	4198	11	AK029048	AK029048 Mus muscu
4	224.5	10.0	594	12	BJ393752	BJ393752 BJ393752
5	218.5	9.7	532	29	TA319G10P	TA319G10P
6	217	9.7	537	14	CD039158	CD039158 T. brucei
C 7	207.5	9.2	2141	13	BQ142519	CD039158 pSHA0051G
C 8	200	8.9	508	28	AQ652212	BQ142519 Contig6 M
9	199.5	8.9	601	12	BJ387574	AQ652212 Sheared D
10	195.5	8.7	1002	29	CNS06D6B	AL393417 T3 end of
11	195	8.7	771	14	CA320325	AL393417 T3 end of
12	191	8.5	633	12	BJ369190	BJ369190 BJ369190
13	190	8.5	718	12	BI750157	BI750157 Fg02_10g0
14	185.5	8.3	574	29	TA315H10P	BI750157 Fg02_10g0
15	182.5	8.1	614	9	AJ273402	AL490202 T. brucei
16	180	8.0	641	12	BJ393925	AJ273402 AJ273402
17	180	8.0	665	13	BQ770462	BJ393925 BJ393925
18	179.5	8.0	3091	11	BC011275	BQ770462 UI-M-FIO-
19	177.5	7.9	650	9	AJ274038	BC011275 Mus muscu
C 20	176	7.8	530	29	CNS010PO	AJ274038 AJ274038
C 21	176	7.8	2121	29	BZ424995	AL153820 Anopheles
22	172.5	7.7	580	9	AJ273745	BZ424995 100023066
23	172.5	7.7	583	9	AJ273947	AJ273745 AJ273745
24	172.5	7.7	593	9	AJ273918	AJ273947 AJ273947
25	172.5	7.7	601	9	AJ273921	AJ273918 AJ273918
26	171.5	7.6	573	14	CA937626	AJ273921 AJ273921
27	170.5	7.6	601	9	AJ273050	CA937626 sav42b10.
C 28	170.5	7.6	1029	29	CNS071DW	AJ273050 AJ273050
29	170	7.6	895	13	BQ216158	AL424794 T7 end of
30	170	7.6	1572	11	AY107161	BQ216158 AGENCOURT
31	169.5	7.5	545	13	BU575479	AY107161 Zea mays
32	169.5	7.5	604	9	AJ273185	BU575479 TgESTZvB8
33	168.5	7.5	601	12	BJ365857	AJ273185 AJ273185
34	168.5	7.5	609	9	AJ272712	BJ365857 BJ365857
35	168.5	7.5	937	10	BG246418	AJ272712 AJ272712
36	168	7.5	1050	29	CNS0780L	BG246418 602360428
C 37	167.5	7.5	564	28	AO651427	AL433387 T3 end of
38	167.5	7.5	610	9	AJ274218	AO651427 Sheared D
39	167	7.4	449	12	BJ359939	AJ274218 AJ274218
40	167	7.4	831	14	CA510555	BJ359939 BJ359939
41	164	7.3	615	9	AJ274059	CA510555 UI-R-FJO-
42	164	7.3	781	13	BQ612128	AJ274059 AJ274059
C 43	163	7.3	586	29	CPA561690	BQ612128 UI-M-EWO-
44	163	7.3	593	9	AJ273903	AJ561690 Cryptospo
45	163	7.3	594	14	CD040823	AJ273903 AJ273903
						CD040823 pSHB0362A

ALIGNMENTS

RESULT 1
BQ622771
LOCUS
DEFINITION BQ622771 1605 bp mRNA linear EST 01-JUL-2002
CC-Contig967 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.
BQ622771
BQ622771.1 GI:21649940
EST:
Conidiobolus coronatus (Delacroixia coronata)
Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)


```
/clone_lib="RIKEN full-length enriched mouse cDNA library"  
/dev stage="10 days neonate"
```

CDS

417. .3575
/note="unnamed protein product; membrane-bound
transcription factor protease, site 1 (MGDI|GI:1927235,
GB|NM_019709, evidence: BLASTN, 99%, match=3782)
putative"
/codon_start=1
/protein_id="BAC26263.1"
/db_xref="GI:26325018"
/translation="MKLVSTMLLVLLVLLCGRLHGLDRLGTRALEKAPCPSCSLHTLK
VEFSTVEVEYIVAGVNFYAKARNSP ISSALKSEVENWRIIPRNPSSDPSDEE
VIQKEKQACGLILDEOHPNLIKTRPQKVFSLKFAESNPVPCNTRNSQKWQSSR
PLKRASLGSQGFHHATGRHSRLLRAIPROVAOTLOADVIALSMRCCQGFAPDAELHIF
TGLSEKHDFHNKRNWTNWTERTLDLGLHGTFVAGVIALSMRCCQGFAPDAELHIF
VFTNQVSYTWFDAFNAILKMDVLNLISGGDFDMDFVDKRWFLTANNLMVLS
AIGNQDPLGYTLNPADQMIVIGQIDFEDNIARLFFSTGMTWELPQSGVGRKVPDI
TVAGVRSGYKGGCRASGTSVASPVVAGAVTLLVSTQKRELVPASVQKALIASA
RLPGVNMPEOGHGKDLDLRAVQILSSYKQPAASLSPSIDLTECPYMPYCSQPIYFG
GMPTLVNVTILNGMGVGRIVDKPWREPYILPQNGNDIEVASVSLWMPWSGYLAISI
SVTKAASWEGIAQHIMITVASPAETELHSGAEHTSTVKILPKVLTPTPPRSKRLV
WDOYHNLRYPCGYFPDRNLRMKNDLGDWDHVNFDMDYHNSMGI FVEVLGAPF
TCFDATQYGTLLVDSEETPEETAKLRDNDNGLSLVIFSDWNTYISVMRKVKFYDE
NRQWMPDGOAGTANTPALNELLSVNMGMFSDGLYEGEFVLHNDHMYYASGSIKAFPE
DGVVLTQTFKQGLVEIKQETA AVENVTIIGLYQIPSEGGGRIVLYGNSCIDDSHRQ
KDCFLLDALLQYTSYGYTTPSLSHSGNQRQPPSGAGLAPPERMEGNHLHRYSKVLEA
HUGDPKRPGLPACPHLSWAKPOPLNATSNLWKHKQKLLSIDLDKAVILPNFRSNPOV
RPLSGESGANDIPGGMGRYINQVEGOTIPVFAELGMAVLAFFVQVLSKAKSRPK
RRPRAKRPLQAOAHPARTPSV"
4180. .4185
polva signal

polyA⁺ signal

polvA_site

BASE COUNT	1016 a	1065 c	1138 g	979 t
ORIGIN				

Alignment Scores:		
Pred. No.:	3.85e-12	Length: 4198
Score:	225.50	Matches: 116
Percent Similarity:	41.32%	Conservative: 65
Best Local Similarity:	26.48%	Mismatches: 160
Query Match:	10.04%	Indels: 97
DB:	11	Gaps: 25

US-09-985-689A-1 (1-434) x AK029048 (1-4198)

Qy	8	VallysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla	27
Db	1005	CTCAGCAGCATGTGCTGG---CAGATGGGATACACAGTCTCTAATGTCAGAGTGTCT	1061
Qy	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47
Db	1062	GTATTTTCATCTGGGCTC-----AGTGAGAAGATCCGCATTATTAAAG	1103

48 GlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63

Db
1104 AAT-----GTGAAGGAGAGACCAACTGGACCAATGAGCGGACCCCTG 1145

QY 64 ---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThr 82

Db
1146 GATGATGGGCTAGGCCATGGCACATTCTGTTGCAGGT---GTGATTGCCAGCATGAGGGAG 1202

QY 83 AsnLysGlyMetAlaProGlnAlaAsnLeu--ValPheGlnSerIleMetAspserGly IUI
 :::||||| ||| :::||||| :::|

Db
1203 TGCCAGGATTTGCTCCAGATGCAGAGCIGACAGATCTTCAGGGGCTTTGCCAGATTT

[illegible]

1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
18

1302 GGCATCCCTAAGAAGATGGACGTTCTCAACCTTAGCATCGGTGGCCCCGACTTCATGGAT 1361

[illegible][illegible][illegible]

QY	346	LysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVal	365
DB	1929	-----ATCTACTATGGAGGATGCCA-----ACAATCGTTAAATGTCCACCATCCTC	1973
QY	366	AsnAspIleuAspLeu-----ValIleThrAlaProAsnGlyThrGlnTyrValGly	382
DB	1974	AATGGCATGGCGTCACAGGAAGAATTGTGCATAAGCTGAGTGCGCACCTATTATA---	2030
QY	383	AsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValcLuAsnVal	402
DB	2031	-----CCACAGCAATGGGA-----GACACCAATTCGAAGTGGCC	2060

QY 403 Phe-----IleAsnAlaProGlnSerGlyThrTyThrIleGluVal 416
 ||| :::: | | | | | | | | |
Db 2061 TTCTCCTACTCCICAGTGTTGTGCCCTGGTCAAGTTACCTTGCATCTCCATT 2114

RESULT 4

[illegible]

discoideum cDNA clone dds32b16 5', mRNA sequence.
ACCESSION RT393752

VERSION BJ393752.1 GI:19304838
KEYWORDS EST.

SOURCE	ORGANISM
Dictyostellium discoideum	Dictyostellium discoideum
Dictyostellium discoideum	Dictyostellium discoideum

REFERENCE
1 (bases 1 to 594)

AUTHORS TITLE JOURNAL COMMENT

Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostellium discoideum at the slug stage
Unpublished
Contact: Tadashi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES Source

1..594
/organism="Dictyostellium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds32b16"
/sex="mat. A"
/dev_stage="Slug stage"
/clone_lib="Dictyostellium discoideum cDNA library, SF"

BASE COUNT 185 a 108 c 107 g 194 t
ORIGIN

Alignment Scores:

Pred. No.: 2,6e-13 Length: 594
Score: 224.50 Matches: 59
Percent Similarity: 50.80% Conservative: 36
Best Local Similarity: 31.55% Mismatches: 59
Query Match: 9.99% Indels: 33
DB: 12 Gaps: 9

US-09-985-689A-1 (1-434) x BJ393752 (1-594)

QY 49 LysileThralaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHis 68
DB 21 AAAGTTTAACTTATATACCATCAACACGACGATAGTATAAAGTGCATGCAC 80
QY 69 GlyThrHisValAlaGlySerValLeuGly-----AsnGlySer 81
DB 81 GTACACATATTTGTGGTCTGCGACGAGTACTCCAGAGGATTTCTTCAGTTAATATTCA 140
QY 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 101
DB 141 TCATTATAGTCTTGCACATGATGCAAGATTCATTC-----TTTGATTGGCA 191
QY 102 GlyGlyLeuGlyGlyLeu-----ProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
DB 192 AGTGGTTCATCAAGTTTGACACCTCCATCGGATTTGAACAATTTATATCAACCATTTAT 251
QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----AlaValAsn 135
DB 252 GACGACGGTCAAGATGCGATTCTGATCTTGGGGTCTGTATCAGTAGAGGGGTATACA 311
QY 136 GlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsn---AspMet 154
DB 312 GTAGTATATTCATCAACACTGCTTCAATTCATGATTTCTCTTTCATCTCATCCAGATTTC 371
QY 155 ThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSer-----Ala 172
DB 372 ATCATCTTAGAGCTGCTGGTAC-----AACGACCAATACCTATCACTACTCACT 422
QY 173 ProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg----- 189
DB 423 CAATCCACTGCAAGAATGTTATTACCGTTGGTGTCTCATCAACAATTCATGAAATTTAT 482
QY 190 -----ProSerPheGlySerTyrAlaAspAsnIleAsn-----His 201
DB 483 TTAAGTATGGTCCAAATATATAAATTAATCAATCATCTGTCGATATAAATCAAGAGTTA 542
QY 202 ValAlaGlnPheSerSerArg 208
DB 543 ATATGTGATTTCGATAGCAGA 563

RESULT 5 TA319G10P

LOCUS
DEFINITION
T. brucei sheared genomic DNA clone 319g10, forward sequence,
genomic survey sequence.
AL492464
VERSION
AL492464.1 GI:11867408
SOURCE
GSS.

ORGANISM

Trypanosoma brucei
Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.

REFERENCE AUTHORS

1 (bases 1 to 532)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE JOURNAL

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhl@sanger.ac.uk

COMMENT

Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUFat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

Source

1..532
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="319g10"

BASE COUNT 112 a 126 c 150 g 143 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 9.4e-13 Length: 532
Score: 218.50 Matches: 61
Percent Similarity: 48.37% Conservative: 28
Best Local Similarity: 33.15% Mismatches: 68
Query Match: 9.72% Indels: 27
DB: 29 Gaps: 6

US-09-985-689A-1 (1-434) x TA319G10P (1-532)

QY 93 ValPheGlnSerIleMetAspSerGlyGlyLeuGly----- 105
DB 2 GTAATGCCAAGATATAATGTTGTCGCCAGGGGCGAAGATTCTTCAGGGGTGGGCTGCC 61
QY 106 -----GlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr 119
DB 62 CATCCCACTCAGAGCTTGCTCCCTCCCGACGAGCTTACTCAAAATTATTCGTCGGGTATAT 121
QY 120 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 139
DB 122 GCGGTGAGCCCGGTGTTCTCAAACTCGTGGGGTTTGTGCTCCCTCCGAGTATTC 181
QY 140 ThrAspSerArgAsnValAspTyrValArgLys---AsnAspMetThrIleLeuPhe 158
DB 182 GCTGTGGAAGGATATGATGATGAGTTTCGAGTAGTTATGACGATCGCTACTTACTTC 241
QY 159 AlaAlaGlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsn 178
DB 242 TCCACTGGCAACAGTATCCAAAGTGGC-----CTAATGACTCCGTGTCGTGTAAGAAC 295

QY	179	AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn	198
Db	296	GTGATGGCGTGGGTGCACACAAACGTTGAGCGCTTCGAAGAC	343
QY	199	IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro	218
Db	344	-----ATTGTTTCTTCGTTTCTTCGTCATGTCACATACAGCGGTAGGATGAACCC	397
QY	219	AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer	238
Db	398	GATCTTTCGTCGCCGGGAAGAGTGTCTGCTCTTCTTCGCAAGCATCAGCT	457
QY	239	SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr	258
Db	458	AAA-----CAATGTAAGTGTGTGGCCAAAGCGGGTTCATCGATGGCAACT	502
QY	259	ProIleValAla 262	
Db	503	GCGCGCGTCCGC 514	
RESULT 6			
CD039158		537 bp mRNA linear EST 09-MAY-2003	
LOCUS		psHA005iG10r.206350 psHA: Soybean host 48 hrs post infection	
DEFINITION		Phytophthora sojae cDNA clone SHA005G10 5, mRNA sequence.	
ACCESSION		CD039158	
VERSION		CD039158.1 GI:30493567	
KEYWORDS		Phytophthora sojae	
SOURCE		Phytophthora sojae	
ORGANISM		Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora	
REFERENCE		1 (bases 1 to 537)	
AUTHORS		Qutob,D., Hraber,P.T., Sobral,B.W.S. and Gijzen,M.	
TITLE		Comparative analysis of expressed sequences in Phytophthora sojae	
JOURNAL		Plant Physiol. 123 (1), 243-254 (2000)	
MEDLINE		20267956	
PUBMED		10806241	
COMMENT		Contact: Tyler B Tyler lab	
FEATURES			
source			
BASE COUNT		102 a 174 c 155 g 104 t	
ORIGIN			
Alignment Scores:			
Pred. No.:	1.37e-12	Length:	537
Score:	217.00	Matches:	69
Percent Similarity:	48.07%	Conservative:	18
Best Local Similarity:	38.12%	Mismatches:	48
Query Match:	9.66%	Indels:	46
DB:	14	Gaps:	8
US-09-985-689A-1 (1-434) x CD039158 (1-537)			
QY	169	ThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeu	188
Db	9	ACGATCAGCAGTCGTCAGGAGCAAGAACGTCATCTCAGTCGCGGCCAGTCTGAACGCC	68
QY	189	ArgProSerPheGlySer-----TyrAlaAspAsnIleAsnHisValAlaGln	204
Db	69	GCGCTTCATTCTCTCCATCCATCCAGTCCGCTGCTGATGAGAACACGGTGGCCTCG	128
QY	205	PheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAlaProGly	224
Db	129	TTCTCTCTCAITGGACCGAGCTCGACGAGCAGAGAGCCGACATCGTAGCCCCGGC	188
QY	225	ThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHis	244
Db	189	ATGTCCATCACCTCTTCGAGTCCGAG-----AACCTGGATCGAC-----ACCAAG	236
QY	245	AspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsn	264
Db	237	TCGTCCGGGTCTGCTCGTCGAGGAACTTCCAGGCTACCCCGTGGTGGCGGTATG	296
QY	265	ValAla-----GlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro	280
Db	297	GCTGTGCTGATCTACGAGTGGTGGTGGTGGAGAAC-----GGCGTG---CCG	350
QY	281	LysPro-----SerLeuLeuLysAlaAlaLeuIleAla	291
Db	351	GACCCACCTACGCGATGATGATCCCGCGCTCTCTGATTAGGCACCTTCTTCTACAC	410
QY	292	Gly-----	292
Db	411	ACTGGCGAGGCGATGTCGCCGCTGTGATCGAGCCCGACCTGGCGTTACGTCGCGTG	470
QY	293	-----AlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGly	308
Db	471	GCGCTGGAGACTGCTGCTAAGACGCTNAACTCTACCCCGACTTCAACACGAGGCTACG	530
QY	309	Arg 309	
Db	531	AAG 533	
RESULT 7			
BQ142519/c		2141 bp mRNA linear EST 24-APR-2002	
LOCUS		Contig6 Metarhizium anisopliae sf. acridum ARSEF 324 Metarhizium	
DEFINITION		anisopliae var. acridum cDNA, mRNA sequence.	
ACCESSION		BQ142519	
VERSION		BQ142519.1 GI:20279578	
KEYWORDS		EST.	
SOURCE		Metarhizium anisopliae var. acridum	
ORGANISM		Metarhizium anisopliae var. acridum	
REFERENCE		1 (bases 1 to 2141)	
AUTHORS		Freimoser,F.M., Screen,S., Baga,S., Hu,G. and St. Leger,R.J.	
TITLE		EST analysis of genes expressed by two different insect pathogenic	
JOURNAL		fungi during optimized secretion of proteins	
COMMENT		Contact: Freimoser F. M. Department of Entomology University of Maryland 4112 Plant Sciences Building, College Park, MD 20742, USA Tel: 301 405 16 13 Fax: 301 314 92 90 Email: ff34@umail.umd.edu.	
FEATURES			
source			
Location/Qualifiers			
1. 2141			
/organism="Metarhizium anisopliae var. acridum"			
/mol_type="mRNA"			
/strain="ARSEF 324"			
/db_xref="taxon:92637"			
/clone_lib="Metarhizium anisopliae sf. acridum ARSEF 324"			
/note="Vector: Unizap; Metarhizium anisopliae sf. acridum			
was grown on insect cuticle and chitin for 24 hours. A			
cDNA library was constructed in the unidirectional Lambda			
vector Unizap."			

```

BASE COUNT      534 a      505 c      631 g      468 t      3 others
ORIGIN

Alignment Scores:
Pred. No.:      1,09e-10      Length:      2141
Score:          207.50      Matches:      73
Percent Similarity: 43.08%      Conservative: 36
Best Local Similarity: 28.85%      Mismatches: 91
Query Match:      9.23%      Indels:      53
DB:              13      Gaps:      12

US-09-985-689A-1 (1-434) x B0142519 (1-2141)
Qy      21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db      1059 GGTGAGGTACTTGGGTATATATATATGACACTGGTGTGACGTCTCC-----1012
Qy      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db      1011 -----CACCCGAGCTTTGGCGGTCCGCCCATCTGGCTCAGGAGCTTCATCAAGGTCAA 958
Qy      61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db      957 AAC---CGTGATGCCACGGCATCGGACTCACTCCGCTGGTACTATT-----GGT 910
Qy      81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp--- 99
Db      909 AGCCGAAGCTACGGTGTGCCAAAATGCCAAGCTCTTTGCTGTCAAGGTCTTGTATGAC 850
Qy      100 ---SerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAla 118
Db      849 CAGGGCAGTGTTCTACTCCGHTATCATCATCGGATGCGAGCTTTGTTGCCAGGACTCC 790
Qy      119 TyrSer-----AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsn 135
Db      789 AAGAGTCGTAACTGCCCAATGGCCACATTCCTCCATGAGTCGGGA-----742
Qy      136 GlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThr 155
Db      741 GGTGGCTACTCGGCTCCGTCACACAGGGTCCCGCTGCTTGGTCAGGTCGTGTCTC 682
Qy      156 IleLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
Db      681 CTTCGCGTCCCGCTGGCAGCATACCCGGATGCCCAACACACTCT---CCCGCTTCC 625
Qy      176 AlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGly 193
Db      624 GAGCCTACTCGCTGCACACTGTTGTGGCACCTCGCTCAGTACACGCCGATCTACCTTTCC 565
Qy      194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp 213
Db      564 AACTAC-----559
Qy      214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
Db      558 GGCAGAGTT---GTCGATATCTTCGCTCCTCGTACCAGGCATTCTCTCCACC-----511
Qy      234 LeuAlaProAspSerSerPheThrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253
Db      510 -----TGGATCAAT-----GGCCGCACCAACACCATCTCTGGC 478
Qy      254 ThrSerMetAlaThrProIleValAlaGlyAsnValAla 266
Db      477 ACCTCCATGGCTACTCCCATATATGCTGCTGTGCTGCC 439

RESULT 8
LOCUS      AQ652212/c
DEFINITION Sheared DNA-8F2.TF Sheared DNA Trypanosoma brucei genomic clone
ACCESSION AQ652212
VERSION    AQ652212.1 GI:5145398
KEYWORDS   GSS.

SOURCE      Trypanosoma brucei
ORGANISM    Trypanosoma brucei
REFERENCE   1 (bases 1 to 508)
AUTHORS     El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
            Fraser,C. and Adams,M.
TITLE       Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
JOURNAL     Unpublished
COMMENT     Other_GSSs: Sheared DNA-8F2.TF
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
            Location/Qualifiers
                location=1..508
                organism="Trypanosoma brucei"
                mol_type="genomic DNA"
                strain="TREU927/4 GUTat 10.1"
                db_xref="taxon:5691"
                clone="Sheared DNA-8F2"
                notes="Vector: pUC18; Site:1: SmaI; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from a cloned population of
                Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                sheared to give a tight size distribution (approx 2 kb).
                The v + i method used for the library construction is
                described in detail in Smith, H.O. and Venter, J.C.
                (Making small insert libraries for whole genome shotgun
                sequencing projects. In Genome Sequencing: A Practical
                Approach, eds. M. Vaudin and B. Borell, Oxford University
                Press, 1999)."
            BASE COUNT      130 a      135 c      129 g      114 t
            ORIGIN

Alignment Scores:
Pred. No.:      7.73e-11      Length:      508
Score:          200.00      Matches:      57
Percent Similarity: 50.30%      Conservative: 28
Best Local Similarity: 33.73%      Mismatches: 56
Query Match:      8.90%      Indels:      28
DB:              28      Gaps:      7

US-09-985-689A-1 (1-434) x AQ652212 (1-508)
Qy      156 IleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThr 175
Db      503 CTTATCTCTCCACGTGGCAACAGTATCCAAAGATGCC-----CTAATGACTCGCTGTCGT 450
Qy      176 AlalysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyr 195
Db      449 GGTAAGAAGCTGATGTGCTGGGGTCACACAAAAGCTG-----TTGACGCTTCG 399
Qy      196 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg 215
Db      398 AAACAC-----ATTGTTCTCTGCTGCTTTTTCGATGCTCAACATACGACGCTAGC 348
Qy      216 IleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAla 235
Db      347 ATGAACCCGATCTTGTGCTGCCGGGAAGAGGTGGTGTGCTCTTCTCTCGCAAA 288

```


363	Db	CAACCTCAATTTTATAATGAAATAATATGGATCATCTCATCAAGGGTCCAACACAT	422
213	QY	AspGlyArgIleIysProAspValMetAlaproglyThrPheIleLeuSerAlaArgSer	232
		: : : : : : : : :	
423	Db	GAITGGTAGAATTGAACCTGATATAGTTGCACCTGGTCAATATATACATGGCAAGATCA	482
233	QY	Ser-----LeuAlaProAspSerSerPheTrpAlaAsnHisAspSer	246
		:::	
483	Db	AATGGTGAGAATTCAACAGACCAATGTGGTGGATGTTCTTTA-----CCAAATGCCAAT	536
247	QY	LysTyrAlaTrpMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla	266
		:::	
537	Db	GGTCTAATGTCAATATCTGTACATCAATGGCAACACCAITGGCAACAGCAGCAACAACA	596
267	QY	GlnLeuArgGluHisPheVal	273
		: : : : : : : :	
597	Db	ATTCTTAGACAATATTTAGTT	617

LOCUS	BI750157	718 bp	linear	EST 25-SEP-2001
DEFINITION	Fg02_10g08_R Fg02_AAFC_ECORC-Fusarium_graminearum mycelium			
ACCESSION	Gibberella zeae cDNA clone Fg02_10g08, mRNA sequence			

ACCESSION	BI750157	
VERSION	BI750157.1	GI:15771959
KEYWORDS	EST.	
SOURCE	Gibberella zeae	
ORGANISM	Gibberella zeae	
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocremycetidae; Hypocreales; Nectriaceae; Gibberella.	
AUTHORS	Harris,L.J., Glassco,T., Rocheleau,H., Allard,S., Chapados,J., Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A, Sprott,D. and Tinker,N.A.	
TITLE	Expressed Sequence Tags from Fusarium graminearum mycelium	
JOURNAL	Unpublished	
COMMENT	Contact: Harris, Linda J. Eastern Cereal and Oilseed Research Centre Agriculture and Agri-food Canada Bldg. 21, Central Experimental Farm, Ottawa, Ontario K1A 0C6, CANADA	

BASE COUNT	147 a	236 c	155 g	164 t	16 others	
ORIGIN						
Alignment Scores:						
Pred. No.:	1.46e-09				Length:	718
Score:	190.00				Matches:	66
Percent Similarity:	44.05%				Conservative:	34
Best Local Similarity:	29.07%				Mismatches:	81
Query Match:	4.46%				Indels:	46
DB:	12				Gaps:	10

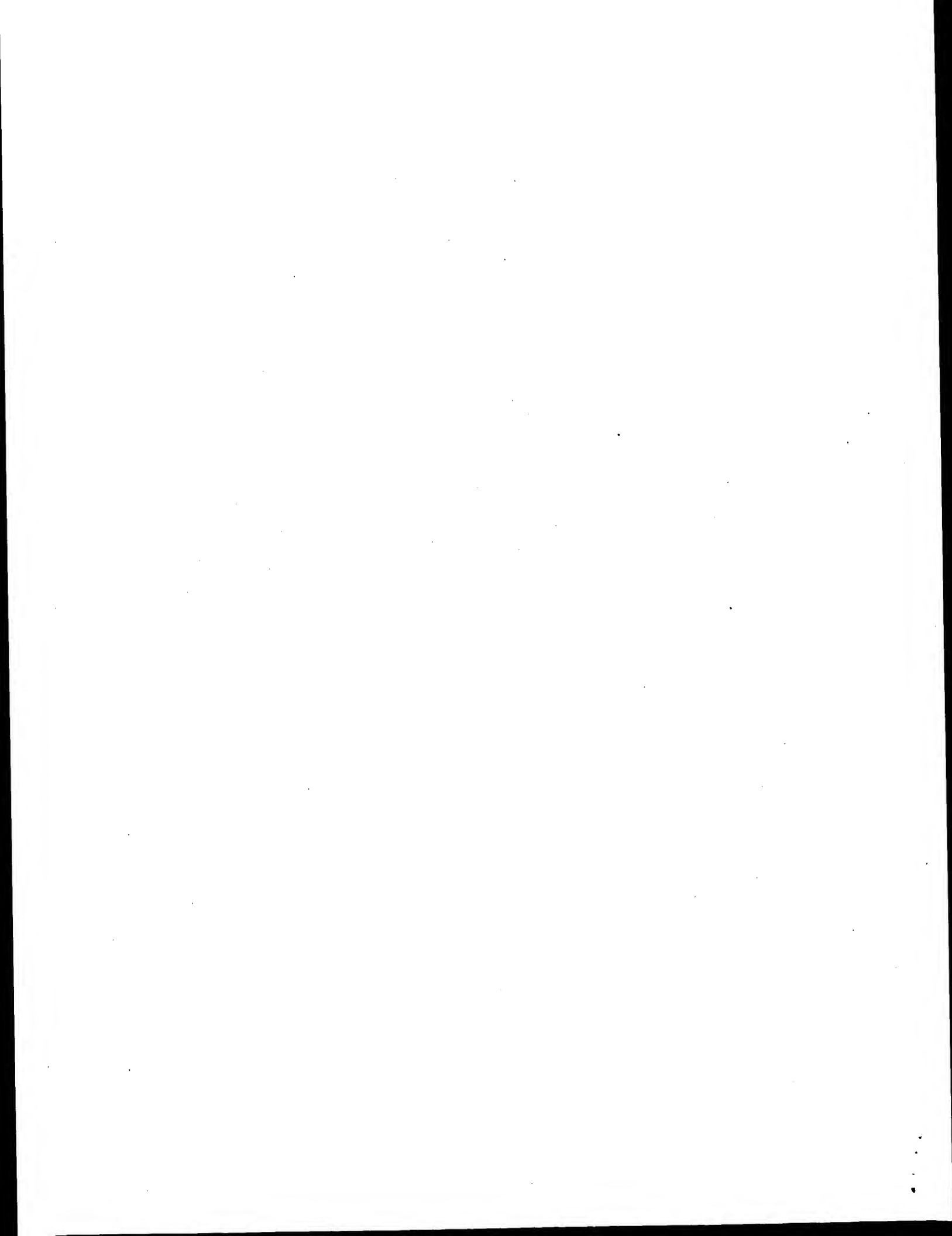
US-09-985-689A-1 (1-434) x BI750157 (1-718)

Qy	46	PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThr	65
Db	9	TTGAGGGTGTGTCTACAGCTGTCTACACTGCCTTCACGGCCCAAGACGCT---GACACC	65
Qy	66	AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnIysGly	85
Db	66	AACGGTACGGAACCTACGTTCTCGCATATATTGCGGAAAG-----ACATACGGT	116
Qy	86	MetAlaProGlnAlaAsnLeu-----ValPheGlnSerIleMetAspSerGly	101
Db	117	GTTCGCAAGAGGCCACCATCCAGCTGTCAAGGTCTTCCAG-----	158
Qy	102	GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer---	120
Db	159	---GGTAGTTTCATCCAGCACCTCCATCATCTCGCTGGCTCAACTGGGCTGCAACGAC	215
Qy	121	-----AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyr	138
Db	216	ATCATCTCCAAGGCCGCAACCAAGACATCAGTCGTCAATATGTCTCGCGCGTGGTAC	275
Qy	139	ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPhe	158
Db	276	TCGTCTTCCTTCAACAACGGTGTGCGAGTCKYTCACGCTCCGGGTATTACTCTGCCATT	335
Qy	159	AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn	178
Db	336	GCTGCCGGAACGATGTGTCACACGCTGCCACACACTTCT---CGTGCTCTCTCCCGAC	392
Qy	179	AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn	198
Db	393	GCCAWCACTCTCGTGCATTGACAGC-----AACTGGGC-----	428
Qy	199	IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro	218
Db	429	-----ATTKCTCGTACTCCACTACGGTAGCGTTCTC-----	461
Qy	219	AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer	238
Db	462	GATATCTTTCGCTGCACTGCAACGACGCTCTCTCCGCG-----	497
Qy	239	SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetClyGlyThrSerMetAlaThr	258
Db	498	-----TGTACACCAAGCAACAGTKCCACCAACCATCAGCGGKACSTCCATGGCTACT	551
Qy	259	ProIleValAlaGlyAsnVal	265
Db	552	CYCCACATTTGCGGACTTGTG	572

LOCUS	TA315H10P	574 bp	DNA	linear	GSS 13-DEC-2000			
DEFINITION	T. brucei sheared genomic DNA clone 315h10, forward sequence, genomic survey sequence.							
ACCESSION	AL490202							
VERSION	AL490202.1	GI:11866292						
KEYWORDS	GSS							
SOURCE	Trypanosoma brucei							
ORGANISM	Trypanosoma brucei							
REFERENCE	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;							
AUTHORS	1 (bases 1 to 574)							
	Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,							
	Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,							
	Meiville, S.E., Rajandream, M.A. and Barrell, B.G.							
TITLE	Direct Submission							
JOURNAL	Submitted (10-DEC-2000)							
	Trypanosoma brucei genome sequencing							
	project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,							
	Cambridge CB30 1SA, E-mail: barrell@sanger.ac.uk and							
	nh@sanger.ac.uk							
COMMENT	Constructed at the Institute for Genomic Research (TIGR).							


```
Db 446 GAGCCTTCTGCTGCACTGTTGGTGCCTCTGCGAAATGACAGCGATCTTCCTCTTC 505
Qy 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp 213
Db 506 AACTAC----- 511
Qy 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
Db 512 GGCAGAGTT---GTCGATATTTCGCTCCTGGTAGCAATGTCTTTCCACC----- 559
Qy 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly 253
Db 560 -----TGGATTGGT-----GGCGGCACAAACACCATCTCTGGT 592
Qy 254 ThrSerMetAlaThrPro 259
Db 593 ACCTNCAATGGCTACTCCC 610
```

Search completed: July 28, 2003, 08:38:57
Job time : 2913.34 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 18:58:19 ; Search time 15.0173 Seconds
(without alignments)
1222.784 Million cell updates/sec

Title: US-09-985-689A-1

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAOSSYGLY.....EVQYNVPEVGFQTFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2.6/ptodata/1/iaa/5A-COMB.pep.*
- 2: /cgn2.6/ptodata/1/iaa/5B-COMB.pep.*
- 3: /cgn2.6/ptodata/1/iaa/6A-COMB.pep.*
- 4: /cgn2.6/ptodata/1/iaa/6B-COMB.pep.*
- 5: /cgn2.6/ptodata/1/iaa/PCTUS-COMB.pep.*
- 6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2247	100.0	640	4	US-09-509-814A-6
2	2242	99.8	640	4	US-09-509-814A-8
3	2183	97.2	639	4	US-09-509-814A-4
4	2155	95.9	639	4	US-09-509-814A-1
5	2155	95.9	640	4	US-09-509-814A-2
6	2125.5	94.6	641	2	US-08-873-479-42
7	1986.5	88.4	433	4	US-09-104-623A-4
8	1986.5	88.4	433	4	US-09-019-532-4
9	1986.5	88.4	433	2	US-08-873-479-43
10	1581.5	70.4	345	4	US-09-512-251A-10
11	1581.5	70.4	345	4	US-09-515-150A-10
12	452.5	20.1	659	3	US-08-894-818B-1
13	452.5	20.1	659	4	US-09-445-472-12
14	414	18.4	412	4	US-09-445-472-1
15	414	18.4	412	4	US-08-894-818B-3
16	414	18.4	522	4	US-09-445-472-4
17	414	18.4	522	4	US-09-445-472-2
18	414	18.4	654	3	US-08-894-818B-35
19	401	17.8	659	3	US-09-445-472-16
20	346	15.4	520	3	US-08-894-818B-5
21	346	15.4	520	3	US-09-000-016-7
22	346	15.4	734	3	US-09-514-340-7
23	346	15.4	734	4	US-09-000-016-4
24	346	15.4	823	3	US-09-514-340-4
25	346	15.4	823	3	US-09-000-016-2
26	304.5	13.6	903	1	US-09-514-340-2
27	304.5	13.6	1398	1	US-08-750-532-1
					US-08-750-532-9

28	304.5	13.6	1398	3	US-08-894-818B-8
29	304.5	13.6	1398	4	US-09-445-472-6
30	295.5	13.2	237	1	US-08-750-532-18
31	283.5	12.6	418	2	US-08-873-479-44
32	276	12.3	418	4	US-09-966-921A-2
33	256.5	11.4	397	4	US-09-328-352-7533
34	254.5	11.3	275	1	US-08-431-387-1
35	251	11.2	280	1	US-08-434-255-8
36	251	11.2	280	1	US-08-459-967-8
37	251	11.2	280	1	US-08-460-327-8
38	251	11.2	280	1	US-08-459-871-8
39	251	11.2	280	3	US-09-024-532-2
40	251	11.2	280	4	US-09-104-623A-2
41	251	11.2	280	4	US-09-019-532-2
42	251	11.2	280	4	US-09-417-359A-2
43	251	11.2	370	1	US-08-434-255-6
44	251	11.2	370	1	US-08-459-967-6
45	251	11.2	370	1	US-08-460-327-6

ALIGNMENTS

RESULT 1
US-09-509-814A-6
; Sequence 6, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIRO
; APPLICANT: KOBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUMI
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 640
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-509-814A-6

Query Match	100.0%	Score	2247	DB 4	Length	640			
Best Local Similarity	100.0%	Pred. No.	2e-173						
Matches	434	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	1	NDVARGIVKADVAQSSVGLYGOGQIVAVADTGLDTRNDSSMHFAFRGKTTALVALGRTN	60						
b	207	NDVARGIVKADVAQSSVGLYGOGQIVAVADTGLDTRNDSSMHFAFRGKTTALVALGRTN	266						
QY	61	NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS	120						
b	267	NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS	326						
QY	121	AGARHTNSWGAAGVNGAYTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI	180						
b	327	AGARHTNSWGAAGVNGAYTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI	386						
QY	181	TVGATENLRPSFGSYADNINHVAFSSRGPTKGRKPKDVMAPCTFTILSARSLAPDSSF	240						
b	387	TVGATENLRPSFGSYADNINHVAFSSRGPTKGRKPKDVMAPCTFTILSARSLAPDSSF	446						
QY	241	WANHDSKAYMGGTSMATPIVAGNVAQLREHFYKNRGITPKPSLLKAAIAGAADICLGY	300						

201405 SID 192
201405
201405

Tue Jul 29 14:12:41 2003

361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 420
 567 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 626

447 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
 301 PNGNOGWRVTLDRKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
 507 PNGNOGWRVTLDRKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 566

361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 420
 567 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 626

421 VPVGPOTFSLAIVN 434
 627 VPVGPOTFSLAIVN 640

RESULT 2
 US-09-509-814A-8

; Sequence 8, Application US/09509814A
 ; Patent No. 6376227
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAIWA, MIKIO
 ; APPLICANT: OKUDA, MITSUYOSHI
 ; APPLICANT: SAEKI, KATSUHIISA
 ; APPLICANT: KUBOTA, HIROMI
 ; APPLICANT: HITOMI, JUN
 ; APPLICANT: KAGEYAMA, YASUSHI
 ; APPLICANT: SHIKATA, SHITSUW
 ; APPLICANT: NOMURA, MASAFUMI
 ; TITLE OF INVENTION: ALKALINE PROTEASE
 ; FILE REFERENCE: 0327-0832-OPCT
 ; CURRENT APPLICATION NUMBER: US/09/509,814A
 ; CURRENT FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: PCT/JP98/04528
 ; PRIOR FILING DATE: 1998-10-07
 ; PRIOR APPLICATION NUMBER: JP 9-274570
 ; PRIOR FILING DATE: 1997-06-08
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 8
 ; LENGTH: 640
 ; TYPE: PRT
 ; ORGANISM: Bacillus sp.
 ; US-09-509-814A-8

Query Match 99.8%; Score 2242; DB 4; Length 640;
 Best Local Similarity 99.8%; Pred. No. 4.9e-173;
 Matches 433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db 207 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
 Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 Db 267 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 326
 Qy 121 AGARHTNSGWAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db 327 AGARHTNSGWAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 386
 Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
 Db 387 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 446
 Qy 241 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 Db 447 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 506
 Qy 301 PNGNOGWRVTLDRKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
 Db 507 PNGNOGWRVTLDRKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 566

Qy 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 420
 Db 567 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 626
 Qy 421 VPVGPOTFSLAIVN 434
 Db 627 VPVGPOTFSLAIVN 640

RESULT 3
 US-09-509-814A-4

; Sequence 4, Application US/09509814A
 ; Patent No. 6376227
 ; GENERAL INFORMATION:
 ; APPLICANT: TAKAIWA, MIKIO
 ; APPLICANT: OKUDA, MITSUYOSHI
 ; APPLICANT: SAEKI, KATSUHIISA
 ; APPLICANT: KUBOTA, HIROMI
 ; APPLICANT: HITOMI, JUN
 ; APPLICANT: KAGEYAMA, YASUSHI
 ; APPLICANT: SHIKATA, SHITSUW
 ; APPLICANT: NOMURA, MASAFUMI
 ; TITLE OF INVENTION: ALKALINE PROTEASE
 ; FILE REFERENCE: 0327-0832-OPCT
 ; CURRENT APPLICATION NUMBER: US/09/509,814A
 ; CURRENT FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: PCT/JP98/04528
 ; PRIOR FILING DATE: 1998-10-07
 ; PRIOR APPLICATION NUMBER: JP 9-274570
 ; PRIOR FILING DATE: 1997-06-08
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 639
 ; TYPE: PRT
 ; ORGANISM: Bacillus sp.
 ; US-09-509-814A-4

Query Match 97.2%; Score 2183; DB 4; Length 639;
 Best Local Similarity 96.3%; Pred. No. 2.9e-168;
 Matches 418; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db 206 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 265
 Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 Db 266 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 325
 Qy 121 AGARHTNSGWAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
 Db 326 AGARHTNSGWAAGVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 385
 Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 240
 Db 386 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTIFLSARSLAPDSSF 445
 Qy 241 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 300
 Db 446 WANHDSKYAYMGTSMTPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAAGADIGLY 505
 Qy 301 PNGNOGWRVTLDRKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 360
 Db 506 PNGNOGWRVTLDRKSLNVAAYVNESSLSSTOKATYSFTATAGKPLKISLVMSDAPASTTA 565
 Qy 361 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 420
 Db 566 SVTLVNDLVLITAPNGTQYVGNDFTSYNDNDGRNNVNFVINAPOSQTYTIEVQAYN 625
 Qy 421 VPVGPOTFSLAIVN 434
 Db 626 VPVGPOTFSLAIVN 639

```
RESULT 4
US-09-509-814A-1
; Sequence 1, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUYOSHI
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (23)..(23)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (29)..(29)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (32)..(32)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (46)..(46)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (47)..(47)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (53)..(53)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (70)..(70)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (74)..(74)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (89)..(89)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (102)..(102)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (105)..(105)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (128)..(128)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (130)..(130)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (131)..(131)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
```

```
; LOCATION: (132)..(132)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (133)..(133)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (146)..(146)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (148)..(148)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (160)..(160)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (165)..(165)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (172)..(172)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (183)..(183)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (187)..(187)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (188)..(188)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (189)..(189)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (194)..(194)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (286)..(286)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (306)..(306)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (324)..(324)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (369)..(369)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (431)..(431)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (501)..(501)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (531)..(531)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (541)..(541)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (584)..(584)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (591)..(591)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (592)..(592)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (594)..(594)
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: (595)..(595)
```

Tue Jul 29 14:12:41 2003

us-09-985-689a-1.ra

OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (611)..(611)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (632)..(632)
OTHER INFORMATION: Xaa is any amino acid

US-09-509-814A-1

Query Match 95.9%; Score 2155; DB 4; Length 639;
Best Local Similarity 96.3%; Pred. No. 5.3e-166;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Caps 0;

QY 1 NDVARGIVKADVAQSSGLYGOGIVAVADTGLDGRNDSSMHEAFRGKITALLYALGRTN 60
Db 206 NDVARGIVKADVAQSSGLYGOGIVAVADTGLDGRNDSSMHEAFRGKITALLYALGRTN 265

QY 61 NANTNGHGTHTVAGSVLNGSTNKGMAPOANLVFQSDSGGLGLPSNLQTLFSQAYS 120
Db 266 NANTNGHGTHTVAGSVLNGSTNKGMAPOANLVFQSDSGGLGLPSNLQTLFSQAYS 325

QY 121 AGARIHTNSWGAANGVAYTDSRNVDDYVRKNDMTILFAAGNEGPNGGTLISAPGTAKNAI 180
Db 326 AGARIHTNSWGAANGVAYTDSRNVDDYVRKNDMTILFAAGNEGPNGGTLISAPGTAKNAI 385

QY 181 TVGATENLRPFSGYADNINHVNAQFSSRGPTKGRIPKPDYMAPGTETILSARSSLAPDSSF 240
Db 386 TVGATENLRPFSGYADNINHVNAQFSSRGPTKGRIPKPDYMAPGTETILSARSSLAPDSSF 445

QY 241 WANHDSKYAYMGTSMTATPIVACNVQAQLREHFVKNRGITPKPSLLKAAALAGADIGLY 300
Db 446 WANHDSKYAYMGTSMTATPIVACNVQAQLREHFVKNRGITPKPSLLKAAALAGADIGLY 505

QY 301 PNGNOGWGRTLDKSLNVAYNVNESSLSSTOKATYFTATAGPKLISLVMSDAPASTTA 360
Db 506 PNGNOGWGRTLDKSLNVAYNVNESSLSSTOKATYFTATAGPKLISLVMSDAPASTTA 565

QY 361 SVTLVNDLVLITAPNGTVGVNDFTSPYNDNDGRNNVENVFNAQSGTYTTEVQAYN 420
Db 566 SVTLVNDLVLITAPNGTVGVNDFTSPYNDNDGRNNVENVFNAQSGTYTTEVQAYN 625

QY 421 VPGVGPOTSLAIVN 434
Db 626 VPGVGPOTSLAIVN 639

RESULT 5
US-09-509-814A-2
Sequence 2, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAEKI, KATSUHIISA
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUW
APPLICANT: NOMURA, MASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-0PCT
CURRENT APPLICATION NUMBER: US/09/509,814A
CURRENT FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JF98/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
PRIOR FILING DATE: 1997-06-08
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2

LENGTH: 640
TYPE: PRT
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3)..(3)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (24)..(24)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (30)..(30)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (33)..(33)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (48)..(48)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (54)..(54)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (71)..(71)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (75)..(75)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (90)..(90)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (103)..(103)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (106)..(106)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (129)..(129)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (134)..(134)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (147)..(147)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (149)..(149)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (161)..(161)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (166)..(166)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (173)..(173)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (184)..(184)
OTHER INFORMATION: Xaa is any amino acid

```
NAME/KEY: misc_feature
LOCATION: (188)..(188)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (190)..(190)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (195)..(195)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (287)..(287)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (307)..(307)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (325)..(325)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (370)..(370)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (432)..(432)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (502)..(502)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (532)..(532)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (542)..(542)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (585)..(585)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (593)..(593)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (595)..(595)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (597)..(597)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (612)..(612)
OTHER INFORMATION: Xaa
NAME/KEY: misc_feature
LOCATION: (633)..(633)
OTHER INFORMATION: Xaa
US-09-509-814A-2
```

```
Query Match 95.9%; Score 2155; DB 4: Length 640;
Best Local Similarity 96.3%; Pred. No. 5.3e-166;
Matches 418; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 NDVARGIVADVAAQSSGYLGQGVAVADTGLTGRNDSSMHEAFKGTALYALGRTN 60
Db 207 NDVARGIVADVAAQSSGYLGQGVAVADTGLTGRNDSSMHEAFKGTALYALGRTN 266
QY 61 NANDTNGHGTAVAGSVLGNSTKNGAPQANLVFQSTMDSGGGLGLPSNLQTLFSOAXS 120
Db 267 NANDTNGHGTAVAGSVLGNSTKNGAPQANLVFQSTMDSGGGLGLPSNLQTLFSOAXS 326
```

```
QY 121 AGARIHTNSWGAANGAYTTDSRVDDYVRKNDMTLLFAAGNECPNGCTISAPGTAKNAI 180
Db 327 AGARIHTNSWGAANGAYTTDSRVDDYVRKNDMTLLFAAGNECPNGCTISAPGTAKNAI 386
QY 181 TVGATENLRPSFGSYADNTNHVAQFSSRGPTKDGRIKPDVMAFGTFLSARSSLAPDSSF 240
Db 387 TVGATENLRPSFGSYADNTNHVAQFSSRGPTKDGRIKPDVMAFGTFLSARSSLAPDSSF 446
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALTAGAADIGLGY 300
Db 447 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALTAGAADIGLGY 506
QY 301 PNGOGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 507 PNGOGWGRVTLDKSLNVAYVNESSLSSTQKATYSFTATAGKPLKISLWSDAPASTTA 566
QY 361 SVTLVNDLGLVTAPNGTQYVGNDFTSFYNDWNGDGRNNVFNINAPQSGTYTIEVOAYN 420
Db 567 SVTLVNDLGLVTAPNGTQYVGNDFTSFYNDWNGDGRNNVFNINAPQSGTYTIEVOAYN 626
QY 421 VPVGPTQFSLAIVN 434
Db 627 VPVGPTQFSLAIVN 640
```

4/6/99

```
RESULT 6
US-08-873-479-42
; Sequence 42, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lynne, Christianson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 641 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-873-479-42
```

```
Query Match 94.6%; Score 2125.5; DB 2: Length 641;
Best Local Similarity 93.5%; Pred. No. 1.3e-163;
Matches 406; Conservative 19; Mismatches 8; Indels 1; Gaps 1;
```

TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
STRAIN: Bacillus sp. Y
US-09-104-623A-4

Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.3e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHAEFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAQNNFGLYGQGIIVAVADTGLDGRNDSMHAEFRGKITALYALGRTN 60
QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 120
DB 61 NDVARGIVKADVAQNNFGLYGQGIIVAVADTGLDGRNDSMHAEFRGKITALYALGRTN 60
QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 120
DB 61 NASDPNGHGTTHVAGSVLNG-ALNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAWN 119
QY 121 AGARIHTNSWGAAGVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
DB 120 AGARIHTNSWGAAPVNGAYTANSROVDEYVRNNDMTVLFAAGNEGPNNGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTFILSARSLAPDSF 240
DB 180 TVGATENLRPSFGSIADNPNHIAQFSSRGPTKDGRIKPDVMAPTFILSARSLAPDSF 239
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLKAALIAAGADIGLY 300
DB 240 WANYNSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLKAALIAAGADIGLY 299
QY 301 PNGNQGWRVTLDKSLNVAIVNESSTLSOKATYSFTATAGKPLKISLWSDAPASTTA 360
DB 300 PSGDQGWGRVTLDKSLNVAIVNEATATGOKATYSFOAQAGKPLKISLWSDAPASTTA 359
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTPSYNDNWDGRNNVNFVINAPOSQGTYTEVOAYN 420
DB 360 SYTLVNDLVLITAPNGKQYVGNDFSPYDNDNWDGRNNVNFVINAPOSQGTYTEVOAYN 419
QY 421 VPVGPQTFSLAIVN 434
DB 420 VPSPQRFSLAIVH 433

RESULT 8
US-09-019-532-4
; Sequence 4, Application US/09019532B.
; Patent No. 6416756
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Prent, Annette
; TITLE OF INVENTION: A Modified Enzyme for Skin Care
; FILE REFERENCE: 4922.204-US
; CURRENT APPLICATION NUMBER: US/09/019,532B
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: 0038/97
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 0754/97
; EARLIER FILING DATE: 1997-06-25
; EARLIER APPLICATION NUMBER: 60/051,381
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: PCT/DK98/000015
; EARLIER FILING DATE: 1998-01-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-019-532-4

Query Match 88.4%; Score 1986.5; DB 4; Length 433;
Best Local Similarity 87.3%; Pred. No. 1.3e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSMHAEFRGKITALYALGRTN 60
DB 209 NDVARGIVKADVAQNNFGLYGQGIIVAVADTGLDGRNDSMHAEFRGKITALYALGRTN 268
QY 61 NANDTNGHGTTHVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 120
DB 269 NANDPNGHGTTHVAGSVLNG-ATNKGMAQANLVFQSIMDSGGGLGGLPSNLQTLFQSAYS 327
QY 121 AGARIHTNSWGAAGVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
DB 328 AGARIHTNSWGAAPVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 387
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTFILSARSLAPDSF 240
DB 388 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTFILSARSLAPDSF 447
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLKAALIAAGADIGLY 300
DB 448 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLKAALIAAGADIGLY 507
QY 301 PNGNQGWRVTLDKSLNVAIVNESSTLSOKATYSFTATAGKPLKISLWSDAPASTTA 360
DB 508 PNGNQGWRVTLDKSLNVAIVNESSTLSOKATYSFTATAGKPLKISLWSDAPASTTA 567
QY 361 SVTLVNDLVLITAPNGTQYVGNDFTPSYNDNWDGRNNVNFVINAPOSQGTYTEVOAYN 420
DB 568 SVTLVNDLVLITAPNGTQYVGNDFTPSYNDNWDGRNNVNFVINAPOSQGTYTEVOAYN 627
QY 421 VPVGPQTFSLAIVN 434
DB 628 VPVGPQTFSLAIVH 641

RESULT 7
US-09-104-623A-4
; Sequence 4, Application US/09104623A
; Patent No. 6303752
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Fatum, Tine Muxoll
; APPLICANT: Deussen, Heinz-Josef
; APPLICANT: Roggen, Erwin Ludo
; TITLE OF INVENTION: A Modified Polypeptide
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; STREET: No. 6303752b No. 6303752disk of No. 6303752th America, Inc.
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/09/104,623A
; FILING DATE: 25-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 5256.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEPHONE: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS:

QY 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 60
QY 61 NANTDNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS 120
Db 61 NASDPNGHGTTHVAGSVLNG-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYN 119
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTANSROVDEYVRNNDMTILFAAGNEGPNGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSIADPNHIAQFSSRGPTKDGRIKPDVMAQGTFFILSARSSILAPDSSF 240
Db 180 TVGATENLRPSFGSIADPNHIAQFSSRGATROGRIKPDVMAQGTFFILSARSSILAPDSSF 239
QY 241 WANHSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGTTPKPSLLKAALIAAGADIGLY 300
Db 240 WANHSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGTTPKPSLLKAALIAAGADIGLY 299
QY 301 PNGNOGWRVTLDKSLNVAQFSSRGPTKDGRIKPDVMAQGTFFILSARSSILAPDSSF 240
Db 300 PNGNOGWRVTLDKSLNVAQFSSRGATROGRIKPDVMAQGTFFILSARSSILAPDSSF 239
QY 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDNNDGRNENNVFINAPQSGTYTIEVQAYN 420
Db 360 SVTLVNDLVLITAPNGTQVGNDFTSYNDNNDGRNENNVFINAPQSGTYTIEVQAYN 419
QY 421 VPVGPQTFSLAIVN 434
Db 420 VPVGPQTFSLAIVH 433

RESULT 9

US-08-873-479-43
; Sequence 43, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lyme, Christianson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear
US-08-873-479-43

Query Match 88.4%; Score 1986.5; DB 2; Length 635;
Best Local Similarity 87.3%; Pred. No. 2.2e-152;
Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 60
Db 203 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 262
QY 61 NANTDNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS 120
Db 263 NASDPNGHGTTHVAGSVLNG-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYN 321
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGTISAPGTAKNAI 180
Db 322 AGARIHTNSGAPVNGAYTANSROVDEYVRNNDMTILFAAGNEGPNGTISAPGTAKNAI 381
QY 181 TVGATENLRPSFGSIADPNHIAQFSSRGPTKDGRIKPDVMAQGTFFILSARSSILAPDSSF 240
Db 382 TVGATENLRPSFGSIADPNHIAQFSSRGATROGRIKPDVMAQGTFFILSARSSILAPDSSF 441
QY 241 WANHSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGTTPKPSLLKAALIAAGADIGLY 300
Db 442 WANHSKYAYMGCTSMATPIVAGNVAQLREHFVKNRGTTPKPSLLKAALIAAGADIGLY 501
QY 301 PNGNOGWRVTLDKSLNVAQFSSRGPTKDGRIKPDVMAQGTFFILSARSSILAPDSSF 360
Db 502 PNGNOGWRVTLDKSLNVAQFSSRGATROGRIKPDVMAQGTFFILSARSSILAPDSSF 561
QY 361 SVTLVNDLVLITAPNGTQVGNDFTSYNDNNDGRNENNVFINAPQSGTYTIEVQAYN 420
Db 562 SVTLVNDLVLITAPNGTQVGNDFTSYNDNNDGRNENNVFINAPQSGTYTIEVQAYN 621
QY 421 VPVGPQTFSLAIVN 434
Db 622 VPVGPQTFSLAIVH 635

RESULT 10

US-09-512-251A-10
; Sequence 10, Application US/09512251A
; Patent No. 6555355
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelson, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349.204-US
; CURRENT APPLICATION NUMBER: US/09/512,251A
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-512-251A-10

Query Match 70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 4.9e-120;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQSSYGLYGQGOIVAVADTGLDTCRNDSSMHEAFRGKITALYALGRTN 88
QY 61 NANTDNGHGTTHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS 120
Db 89 NANTDNGHGTTHVAGSVLNG-ATNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYS 147

QY 121 AGARIHTNSWGAAYNGAYTTDSNRVDYVYRKNDMTILFAAGNEGPGGTTISAPGTAKNAI 180
DB 148 AGARIHTNSWGAAYNGAYTTDSNRVDYVYRKNDMTILFAAGNEGPGGTTISAPGTAKNAI 207
QY 181 TVGATENLRPSGSGYADNINHVAFSSRGPTKGRKIPDYMAGPCTFILSARSSSLAPDSSF 240
DB 208 TVGATENLRPSGSGYADNINHVAFSSRGPTKGRKIPDYMAGPCTFILSARSSSLAPDSSF 267
QY 241 WANHDSKYAYMGTSMTATPIVAGNVQALREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
DB 268 WANHDSKYAYMGTSMTATPIVAGNVQALREHFVKNRGITPKPSLLKAAIAGAADIGLGY 327
QY 301 PNGNOGWGRVTLDKSLNV 318
DB 328 PNGNOGWGRVTLDKSLNV 345

RESULT 11
US-09-515-150A-10
; Sequence 10, Application US/09515150A
; Patent No. 6558938
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5348.204-US
; CURRENT APPLICATION NUMBER: US/09/515.150A
; CURRENT FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-515-150A-10

Query Match 70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 4.9e-120;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 29 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 88
QY 61 NANTNGHGTIVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGSLPNSLOTLSOAYS 120
DB 89 NANTNGHGTIVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGSLPNSLOTLSOAYS 147
QY 121 AGARIHTNSWGAAYNGAYTTDSNRVDYVYRKNDMTILFAAGNEGPGGTTISAPGTAKNAI 180
DB 148 AGARIHTNSWGAAYNGAYTTDSNRVDYVYRKNDMTILFAAGNEGPGGTTISAPGTAKNAI 207
QY 181 TVGATENLRPSGSGYADNINHVAFSSRGPTKGRKIPDYMAGPCTFILSARSSSLAPDSSF 240
DB 208 TVGATENLRPSGSGYADNINHVAFSSRGPTKGRKIPDYMAGPCTFILSARSSSLAPDSSF 267
QY 241 WANHDSKYAYMGTSMTATPIVAGNVQALREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
DB 268 WANHDSKYAYMGTSMTATPIVAGNVQALREHFVKNRGITPKPSLLKAAIAGAADIGLGY 327
QY 301 PNGNOGWGRVTLDKSLNV 318
DB 328 PNGNOGWGRVTLDKSLNV 345

RESULT 12
US-08-894-818B-1
; Sequence 1, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 659 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-894-818B-1

Query Match 20.1%; Score 452.5; DB 3; Length 659;
Best Local Similarity 30.1%; Pred. No. 2e-28;
Matches 138; Conservative 67; Mismatches 153; Indels 101; Gaps 18;
QY 8 VKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALY-ALGRTNNANDTN 66
DB 145 ICADTVNNSLGGVDSGVVAIVDTGIDAN-----HPDLKGVIGWYDAVNGRSTPYDDQ 198
QY 67 GHGTHVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGSLPNSLOTLSOAYS 121
DB 199 GHGTHVAGSVLNGSTNKGMAQANLVFQSIMDSGGGLGSLPNSLOTLSOAYS 258
QY 122 GARI-----HTNSWGAAYNGAYTTDSNRVDYVYRKNDMTILFAAGNEGPGGTT 169
DB 259 GIRVINLSLGGSSQSDGDTLSQAVNNANDA-----GIVCVAAAGSGNFTYT 306
QY 170 ISAPGTAKNAITVAGTENRPFSGSYADNINHVAFSSRGPTKGRKIPDYMAGPCTFILS 229
DB 307 VGSPPAAASKVITVGA-----VDSNDNIASFSSRGPTADGRLKPEVAVPGVDIIA 355
QY 230 ARSSLAPDSSFWANHDSKYAYMGTSMTATPIVAG-NVAQALREHFVKNRGITPK--PSLLK 286
DB 356 PRAS---GTSMTGTPINDYTKASGTSMTATPIVAGSVGVALILQAH-----PSWTPDKVK 404
QY 287 AALIAGA-----ADIGLGPNGNOGWGRVTLDKSL---NVAYVNESSSLSSTOKATY 335
DB 405 TALITETADIVAPKEIADIAYGA-----GRVNVYKAIKYDDYAKLITFTGVSADKGSATH 457

Db 370 FEKVGYNPTDGTWIKVVSYS---GSANTQVDVVS 402
Search completed: July 25, 2003, 19:02:38
Job time : 17.0173 secs

CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
APPLICATION DATA: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 522 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
OTHER INFORMATION: /note= Xaa at position 428 is Gly or Val.
US-08-894-818B-3

Query Match 18.4%; Score 414; DB 3; Length 522;
Best Local Similarity 29.8%; Pred. No. 1.8e-25;
Matches 136; Conservative 59; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLVGQGVAVADTGLDTRNDSSMHEAFRCITALLYALGRNTNAN-----DTNGHGTH 71
Db 22 GYDGSGITIGIITGID-----ASHPDLOGKV-----IGWVDFVNGRSYPYDDHGHGTH 70
QY 72 VAGSVLNGSTN---KGMAPQANLVQSIM--DSGGGLGLPSNLQTLFSQAYSAGARI 125
Db 71 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISTIIKGVWAVDNKDKYGIKV 130
QY 126 HTNSWGA-----AVNGAYTTDSRVNDDYVRKNDMTILFAAGNEGPNGGTISAP 173
Db 131 INLSGSSQSSDGTALSQAVNAAWDA-----GLVVVVAAGNSGPNKYTICSP 178
QY 174 GTAKNATVIGATENLRFSGSYADNINHVAQFSSRGPTKDGRIKPDVMAPGTFFLSARSS 233
Db 179 AAASKVITVGA-----VDKYDVITTSFSSRGPTADGRKPEVAVPAGNWIILAAARAS 227
QY 234 LAPDSSFWANHDSKYANGTSMATPIVAGNVAQLREHFVKNRGITPK--PSLLKAALIA 291
Db 228 -----GTSMGOPINDYTTAAPTSMATPHVAGTAALLQ-----AHPSTPKVKITALIE 277
QY 292 GA-----ADIGLGPNGQGWGRVTLDSLNVAIVYVNESSLSSTSQKA-----TYSFT 338
Db 278 TADIVKPEDEADIAGA-----GRVNAVYKAIN--YDNYAKLVFTGYVANKGSOHQHFV 328
QY 339 ATAGKPLKISLVWSDAPASTTASVTVLNDLVLITAPNGTQVGNDFTSFYNDNDWDCRNN 398
Db 329 ISGASFVTATLYDNAN-----SDLDLYLDPNQNGQ--VDYSYTAIY-----G 369
QY 399 VENVFNAPOSGTYTIEVQAYNVVPVGPQTFFSLAIVN 434

220 230 240 250 260 270 280
 KPDYMAPGTFILGARSRLAPDSFVWANHDSKYAYMGGTSNATPIVAGNVAQUREHFVNKRGITPKPSLLKAA
 KPDYMAPGTFILGARSRLAPDSFVWANHDSKYAYMGGTSNATPIVAGNVAQUREHFVNKRGITPKPSLLKAA
 KPDYMAPGTFILGARSRLAPDSFVWANHDSKYAYMGGTSNATPIVAGNVAQUREHFVNKRGITPKPSLLKAA
 220 230 240 250 260 270 280

$$x_{VN-VH}^x$$